OXFORD

趨向空無

TOWARDS NON-BEING

the logic and metaphysics of intentionality

GRAHAM PRIEST

TOWARDS NON-BEING

This page intentionally left blank

Towards Non-Being

The Logic and Metaphysics of Intentionality

Graham Priest

CLARENDON PRESS • OXFORD

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2005

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 0–19–926254–3 978–0–19–926254–0

1 3 5 7 9 10 8 6 4 2

For Richard (1936–1996)—non-existent object though you may now be, your Sosein *is still with us.*

Preface

'If meinongianism isn't dead, nothing is,' Gilbert Ryle is reputed to have said in the heyday of Oxford Philosophy.[1] I think that Ryle was exactly right. No idea in philosophy is ever past its use-by date, at least, no idea of any substance. We may always come back and find new depths in it, new applications for it, new answers to objections that were taken to be decisive. Thus, for example, platonism has re-emerged many times in the history of Western philosophy, most recently in a perhaps unexpected place: in connection with technical results in the foundations of mathematics. Aristotelian virtue ethics has reappeared recently after several hundred years of the dominance of ethics by kantianism and utilitarianism. And so the list goes on.

Of course, this is not how Ryle intended his words to be understood. What he meant was that meinongianism was dead for all time. It would perform no Lazarus-like return. For many years I shared Ryle's view. Educated about thirty years ago in Britain, I took it for granted that Russell had shown that meinongianism was little more than superstition (though one that he himself had subscribed to for some time), and that Quine had shown that it was all just simple obfuscation. That which exists is that over which one can quantify; and that's that.

Thus it was that I was outraged when I met Richard Routley (Sylvan as he later became) in the mid-1970s, and found him stoutly defending a version of meinongianism. (Richard never defended a view in any other way!) I could not understand how the view could possibly be taken seriously. It was my good fortune not just to have met Richard, but to have been able to talk with him about the matter over many years. He persuaded me that all the knock-down arguments that I thought I had were lame or just begged the question; that meinongianism is a very simple, natural, and common-sense view; that the theory has many applications to areas of philosophy where more orthodox views creak at the joints.

[1] I have not been able to track down the source of this quote, so it may just be hearsay. The nearest I have found is '*Gegenstandstheorie* . . . is dead, buried and not going to be resurrected,' Ryle (1973), 255.

Part of the beauty of meinongianism—or at least of Richard's approach to it, spelled out at length in *Exploring Meinong's Jungle* (1980)—is its technical simplicity. To do full justice to the idea one would seem to need impossible worlds, but these one has anyway, at least if one subscribes to some version of relevant logic. But the main technical trick is just thinking of one's quantifiers as existentially neutral. '$\forall$' is understood as 'for every'; '$\exists$' is understood as 'for some'. Existential commitment, when required, has to be provided explicitly, by way of an existence predicate, E, which, *pace* the way that Kant is often—and erroneously—interpreted, is a perfectly normal predicate. Thus, 'there exists something such that' is $\exists x(Ex \wedge \ldots x \ldots)$; and 'all existing things are such that' is $\forall x(Ex \supset \ldots x \ldots)$. The action of the theory is mainly, therefore, not at the technical level, but at the philosophical level.

Indeed, in one way, Richard's view was even simpler than Meinong's. Meinong held that concrete objects exist, abstract objects, like numbers and propositions, subsist, and that merely possible and impossible objects do not in any way exist. They neither exist nor subsist—as Meinong is frequently represented as having claimed. Richard simplified this: concrete objects exist; everything else (abstract objects, worlds, merely possible objects, impossible objects) simply do not exist. (Indeed, though I shall not follow him in this, Richard held the even stronger view that it is only present concrete objects that exist. Past and future concrete objects have the same existential status as abstract objects—none.) To distinguish his view from Meinong's, Richard coined the neologism *noneism* for it, a usage that I will follow.

There is one technical problem that was never really solved in the 1,000-odd pages of *Exploring Meinong's Jungle*, however. This is the characterization problem. Meinong insisted that the *Sein* (being) of an object is independent of its *Sosein* (properties). In particular, objects can be characterized in various ways, and have the properties that they are characterized as having, whether or not they exist—existential status is irrelevant. Thus, we specify an object by a certain set of conditions. These might be: *was a detective*, *lived in Baker St.*, *had unusual powers of observation and inference*, etc. Let us write the conjunction of these conditions as $A(x)$. Then if we call the object so characterized 'Sherlock Holmes', s for short, then s has its characterizing properties, $A(s)$, plus whatever properties follow from these. The idea that an object has those properties that it is characterized as having is called the *characterization principle* (CP). It explains, amongst other things, how we can know some of the things we do about non-existent objects: we know that objects characterized in certain ways have those properties, precisely because they *are* characterized in that way.

Now, the trouble with this idea is that the CP cannot be correct in full generality. If it were, not only could one run the ontological argument to prove the existence of God—and everything else—one could, in fact, prove everything. For let B be any sentence, and consider the condition $x = x \wedge B$. Let t be the object characterized by this condition. Then the CP gives us: $t = t \wedge B$, from which B follows. It would seem, then, that only a restricted class of contexts, $A(x)$, can be used in the CP. The problem is, which? This is the characterization problem. There are various gestures towards a solution to the problem in *Exploring Meinong's Jungle*, but Richard never achieved there—or anywhere else as far as I am aware—a solution that he regarded as fully satisfactory.

I first became attracted to noneism when I found an approach to the characterization problem that I found plausible. The CP *can* hold unrestrictedly, provided only that its instances may hold, not at this world, but at others. This solution is explained in detail in Ch. 4. That chapter began life as a paper given at a conference in St Andrews in 1999. In later discussions, Byeong Yi persuaded me that I had not done justice to the indeterminacy of some intentional states. Chapter 3, which is an attempt to do better, arose out of discussions with Steven Read. Chapters 3 and 4 are a statement of noneism—or my version of it anyway.

In 2001 Jay Garfield invited me to give the annual Alice Ambrose Lazerowitz and Thomas Tymoszko Memorial Logic Lecture at Smith College. I decided to talk about some problems in epistemic logic, and especially the behaviour of identity in epistemic contexts. The first two chapters of the book expand on the topic of that lecture. With this material in place, it became clear that a coherent semantic and metaphysical picture of intentionality in general was available. This is what is presented in Part 1 of the book (Chs. 1–4).

Part 2 of the book (Chs. 5–8) comprises a defence of the view, and especially its noneism, against some natural objections. In the process, a noneist account of various kinds of objects other than the objects of intentional states is articulated. It is widely thought that Quine had demolished meinongianism in his 'On What There Is' (1948); but Quine's demolition was itself demolished by Routley in 'On What There Isn't' (1982). Since the authority of Quine's paper is still widely appealed to, it seems worth repeating Richard's critique. This is essentially what Ch. 5 does. Chapter 6 makes the obvious application of noneism to fiction, and considers various objections to the application; and Ch. 7 applies it to mathematical objects and to worlds. This chapter also picks up a variety of further criticisms, including those of David Lewis's 'Noneism or Allism' (1990).

What, in fact, I take to be the hardest objection to noneism is none of the above. Noneism is naturally committed to the idea that every term denotes something. If one also subscribes to the naive, unrestricted principles that govern semantic notions—and especially denotation (as Richard did, and I do)—then quite unacceptable consequences appear to follow. The argument that shows this was brought to my attention (though not in the context of noneism) by Uwe Petersen in the early 1990s. I discussed the argument with Richard on a couple of occasions before his death in 1996. The discussions were inconclusive. At that time I was inclined to reject the claim that every term denotes. Noneism requires a different answer. Chapter 8 is my attempt to give one.

I presented a version of all this material in a series of seminars in St Andrews towards the end of 2001. At the end of the year Peter Momtchiloff at Oxford University Press suggested that it might be worked up into a short monograph, a suggestion for which I am very grateful. The present book is the result.

It does not attempt to be a comprehensive book on the subjects with which it deals. There are many contemporary approaches to intentionality;[2] several writers other than Richard have defended versions of meinongianism;[3] and many writers have discussed fictional objects and their status.[4] The views of several of these writers share points in common with the approach presented here, as well as important differences. Though I make occasional comments about some of these views when it seems useful to do so, I shall, for the most part, say nothing about them—which should not be taken to imply that I think them unimportant. (Perhaps this would have been a better book had I included comprehensive discussions of all the contemporary accounts; it would certainly have been one that was a lot longer!) A fortiori, I make no systematic attempt to argue that the approach I give here is better than these—though I do

[2] Some idea of the range can be found by consulting the papers in Salmon and Soames (1988), and Anderson and Owens (1990).

[3] e.g. Fine (1982), Parsons (1980), Zalta (1988).

[4] To name but a few, Currie (1990), Lewis (1978), Walton (1990). A survey can be found in Howell (1998).

think that it has a simplicity and directness that makes many alternatives look contrived.[5] Nor do I take the version of the view presented here to be definitive. A number of the techniques developed in the book are relatively novel and untried, and I would be surprised, indeed, if better techniques could not sometimes be found. Finally, it is not my aim here to present a comprehensive defence of the view in question. Though I take up what seem to me to be many of the most important questions and objections, I am sure that there are numerous others, probably even important others, that are not addressed. My aim in the book is to do none of these things. It is very modest: simply to put into play a certain view, to present what seems to me (at least at present) to be the most viable version of a noneist account of intentionality. I shall feel that I have succeeded if the account is attractive enough to merit consideration, and robust enough, generally speaking, to stand up for itself.

At various points in the book there is relevant material that is not essential to the central development of the book. Though this certainly adds to the picture in important ways, it may be skipped without endangering an understanding of the rest. I have therefore put it into appendices for the various chapters. The material is of varied kinds. The appendix to Ch. 6 is a short story. The appendix to Ch. 3 is a discussion of medieval accounts of intentionality. Perhaps most importantly, the appendices to Chs. 1, 2, 4, and 8 present the proofs of the various technical results of a logical kind that are reported in those chapters.

This is not a book on formal logic; but it does deploy techniques of formal logic—necessarily so, since it is a book about formal semantics. As so often, the fact that a metaphysical view has a rigorous logical underpinning, gives it both a precision and a viability that it would not otherwise enjoy. The book also starts by throwing non-logicians in at the deep end—well, not the shallow end, anyway. Much of the important technical material in the book is covered in the first two chapters. Some may heave a sigh of relief at the end of these. It is necessary to proceed in this way, though. You can't discuss the philosophical issues that *X* raises until you have some understanding of *X*.

I have made no attempt to explain those parts of logic that are relatively standard. Thus, I presuppose familiarity with the basics of first-order logic. Any first course in formal logic should be an adequate background. A familiarity with the basics of quantified modal logic would also be an

[5] Much criticism of opposing views is given by Richard in *Exploring Meinong's Jungle*.

advantage. Fitting and Mendelsohn (1998), chs. 1–4, could be consulted for an appropriate exegesis. Although the book deploys the techniques of logical semantics, it does not engage with formal proof procedures, such as tableau methods; nor, therefore, with issues of completeness, etc. That is appropriate material for another study; but not this one. As far as notation goes, since I would like the book to be accessible to non-logicians in so far as is possible, I have opted for readability rather than rigour. Much of the notation is standard. Where it is not, I explain.

The book draws on some material that has already been published—though I should warn that it has been reshaped in many places in the development of the book. In particular, I have used material from 'The Hooded Man', *Journal of Philosophical Logic* (2002), 'Intentionality—Meinongianism and the Medievals', *Australasian Journal of Philosophy* (2004) (written with Stephen Read); 'Objects of Thought', *Australasian Journal of Philosophy* (2000*a*); 'Sylvan's Box', *Notre Dame Journal of Formal Logic* (1997*a*); and 'Meinong and the Philosophy of Mathematics', *Philosophia Mathematica* (2003). I am grateful to the editors of those journals for permission to reuse the material.

Finally, some acknowledgments. As is clear, my biggest debt here is to Richard. Indeed, the book is a small attempt to repay Richard for his intellectual stimulation over some twenty years. To Stephen Read, I owe a very specific debt: he is the co-author of Ch. 3. More general debts, that are harder to define, are owed to many of my colleagues at the Universities of Queensland, Melbourne, and St Andrews, where I worked throughout the development of the material, as well as those who made helpful comments when versions of papers relevant to the book were given at various universities in Australia, the United Kingdom, China, Taiwan, Japan, and the USA—or sometimes in correspondence. These include Max Deutsch, Laurence Goldstein, Allen Hazen, Jesper Kallestrup, Arnie Koslow, Joe Lau, David Lewis, Fraser Macbride, Daniel Nolan, Calvin Normore, Roy Perrett, Uwe Petersen, Agustin Rayo, Stephen Read, Greg Restall, Shibata Masoyoshi, John Skorupski, Barry Taylor, Achille Varzi, Wen-Fang Wang, Crispin Wright, Ed Zalta—and doubtless others whom I have forgotten. I thank all these people warmly. Thanking is, of course, an intentional relation. So on with the story.

Contents

Detailed Table of Contents

PART I SEMANTICS FOR INTENTIONALITY

This page intentionally left blank

Part I

Semantics for Intentionality

This page intentionally left blank

‘Even non-existents can be signified by a name.’

Aristotle (*Posterior Analytics* 92^b29–30).

1
Intentional Operators

1.1 Introduction: Intentionality

Intentionality is a fundamental feature—perhaps the fundamental feature—of cognition. Intentionality is that feature of a mental state whereby it is 'directed towards' an object of some kind. It is recorded linguistically in verbs such as 'know', 'believe', 'fear', 'worship', 'hope', and so on. The semantics of these locutions, with its attendant metaphysical picture, is what this book is about.

The foregoing words hardly provide a definition of intentionality, but neither is one necessary for the present enterprise: an intuitive understanding is quite sufficient.[1] Brentano took intentionality to be the definitive characteristic of a mental state.[2] This might certainly be disputed. It is not obvious, for example, that being in pain or generally feeling anxious are intentional.[3] But that matter does not need to be resolved for present purposes, either. More important for our purposes is that intentional verbs are notorious for generating puzzles and conundrums. Many of these will concern us in due course.

Intentionality, and the problems to which it gives rise, have been discussed in Western philosophy since Ancient Greece. Thus, for example, problems about substitutivity of identicals are discussed by Aristotle in *De Sophisticis Elenchis* (179^a24–179^b34). Problems about the analysis of intentional contexts are discussed by the great medieval logicians, such as Ockham and Buridan.[4] Intentionality plays centre-stage in the work

[1] For a general introduction to the notion of intentionality, see Crane (1998).

[2] See ch. 1 of Chisholm (1960).

[3] For some discussion of the issue, see Searle (1983), 1–2, and Crane (1995), 37–40.

[4] For a general discussion of epistemic logic in the later Middle Ages, see Boh (1993). For more detailed references to Ockham, Buridan, and other medieval logicians, see 3.7.

of Brentano[5] and his erstwhile disciple Meinong. It is discussed by Frege and the early Russell.[6] In the first half of the twentieth century the notion took rather a hammering, however. It is somewhat cursorily dismissed by Wittgenstein in the *Tractatus* (5.541–2), and Quine launched a swingeing attack on it some thirty years later.[7] The fortunes of intentionality revived with the appearance of possible-world semantics in the 1960s, when a number of people, such as Hintikka,[8] realized that these might be applied to the analysis of intentional contexts. World-semantics for intentionality have also occasioned some further philosophical discussion, e.g. by Kripke.[9]

But despite the recent renewed interest in intentional contexts, the semantics of intentionality are in a highly unsatisfactory state. There are well-known possible-world semantics for some intentional notions, notably knowledge and belief; but these, as we will see, are subject to also well-known problems, such as the problem of logical omniscience and issues concerning the substitution of identicals. Contemporary discussion of how to treat intentional verbs with non-propositional objects is all but non-existent, and the status of possible (and impossible) objects is an ongoing debate. In the chapters that follow, I will attempt to provide a coherent analysis of intentional contexts. This will involve a semantics of intentional verbs, the logical principles concerning intentional notions that the semantics generate, and, above all, the metaphysical interpretation of the semantics.

1.2 Operators and Predicates

Let us start by observing that intentional verbs can take different kinds of complements. Sometimes the grammatical complement of an intentional verb is a noun-phrase, as in:

> Ponce de Leon sought the Philosopher's Stone.
>
> The Ancient Greeks worshipped Zeus.
>
> George W. Bush fears an attack by Osama Bin Laden.

[5] Brentano (1874); see the first three chapters of Chisholm (1960). For the medieval genesis of Brantano's notion of intentionality, see Sorabji (1991).

[6] For references to Frege, see 2.7, and to Russell and Meinong, see Ch. 5.

[7] e.g. Quine (1948, 1956).

[8] Hintikka (1962, 1969).

[9] As we will see in 2.6.

The complement can, however, also be a sentence (possibly preceded by 'that'), as in:

> John Howard believes (that) he is a great prime minister.
>
> George W. Bush dreamed (that) he was a lizard.
>
> Kofi Annan fears (that) the invasion of Iraq has destabilized the Middle East.

The object towards which the mental state is directed in cases of this kind can be thought of as the proposition expressed by the embedded sentence. Note that intentional complements of this kind can also appear in an accusative-infinitive form, as in: John Howard believes himself to be a great prime minister.

I will call intentional verbs with noun-phrase complements *predicates* and intentional verbs with sentential complements *operators*.[10] Some verbs can be only predicates:

> I worship you.

is okay; but:

> *I worship that you like me.

is not. Some verbs can be only operators:

> I dreamed that you love me.

is okay: but:

> *I dreamed you.

is not. Some can be both:

> I fear you.
>
> I fear that you will go.

are both okay. There seems to be no systematic connection between a verb used as a predicate and the same verb used as an operator[11]—or at

[10] Note that intentional predicates are normally binary; but there are certainly intentional predicates of greater adicity. For example: x prefers y to z.

[11] Thus, one might suggest that if v is a verb that can be both, x vs y iff for some A, x vs that $A(y)$. But this will not work in general. If I have forgotten you then doubtless there are many things about you that I have forgotten. But if I have forgotten that you dislike Wagner, it does not follow that I have forgotten you.

least, if there is, it will not concern us here. So I will treat a verb that can be used in both ways as ambiguous. Thus, e.g. I will take the operator '. . . fears that . . .' and the predicate '. . . fears . . .' to be distinct.

The semantics of intentional operators and those of intentional predicates are closely connected, as we shall see in due course; but our concern in this chapter and the next will be with operators. We will turn our attention to predicates in the subsequent two chapters. By the end of Part 1 of the book, we will have an integrated account for both kinds.

It is worth noting that these two sorts of context do not exhaust all intentional contexts. An intentional verb can also be followed by various other constructions, as the following demonstrate:

> The IRA know where to get nuclear weapons.
>
> Nobody knows who Jack the Ripper was.
>
> John Howard knows how to scare the Australian voters.

I shall make comments on such constructions occasionally, but by and large I shall ignore them. Complements of the kind in question would appear to be closer to propositional complements than nominal complements, but they also appear to be very specific to just a few intentional verbs—notably, 'know'. Thus, none of the following makes sense, for example:

> *The IRA desires where to get nuclear weapons.
>
> *Nobody hopes who Jack the Ripper was.
>
> *John Howard fears how to scare the Australian voters.

Particular intentional verbs may certainly have distinctive properties of their own. I will comment on these from time to time, but I will not discuss individual verbs at any length. It is the semantics of the general class of intentional verbs that is my primary focus.

1.3 World Semantics

So what should a semantics for a language that contains intentional operators look like? To the extent that there is currently any orthodoxy on the matter, the answer is that provided by some kind of world semantics. Let me start by spelling out a form of this.

Take an ordinary first-order language. It has a set of constants, n-place function symbols, and n-place predicates, including 0-place predicates (propositional parameters). Until the next chapter, we will assume that it does not contain identity. Augment this language with a collection of intentional operators. I will write these as upper case Greek letters. Thus, if t is any term and A is any formula, $t\Psi A$ is a formula (t Ψs that A). Though modal operators will not play much role in our discussions, it will be useful occasionally to be able to appeal to these. So we will also assume that the language has the usual modal operators, $\Box$ and $\Diamond$.

An interpretation, $\mathfrak{I}$, for the language is a structure, $\langle \mathcal{C}, @, D, \delta \rangle$. $\mathcal{C}$ is a set of worlds. Each is closed under entailment, so we will call them *closed worlds*. $@ \in \mathcal{C}$ is the actual world. D is a non-empty domain of objects, and δ assigns every non-logical symbol a denotation, thus:

if c is a constant, $\delta(c) \in D$

if f is an n-place function, $\delta(f)$ is an n-place function on D

if P is an n-place predicate, and $w \in \mathcal{C}$, $\delta(P, w)$ is a pair, which I will write as $\langle \delta^+(P, w), \delta^-(P, w) \rangle$

if Ψ is any intentional verb, $\delta(\Psi)$ is a function that maps each $d \in D$, to a binary relation on $\mathcal{C}$. I will write $\delta(\Psi)(d)$ as R^d_Ψ

The first two clauses are orthodox. Let me say a little about the other two.

Let D^n be the set of n-tuples of members of D, $\{\langle d_1, \ldots, d_n \rangle : d_1, \ldots, d_n \in D\}$. Note that, by definition, $\langle d \rangle$ is just d. D^0 is not usually defined; but for the sake of uniformity, we define it as $\{\langle\rangle\}$, where $\langle\rangle$ is the empty sequence. If P is an n-place predicate, $\delta^+(P, w), \delta^-(P, w) \subseteq D^n$; $\delta^+(P, w)$ is the *extension* of P at w, and $\delta^-(P, w)$ is its *co-extension*. Intuitively, the extension of an n-place predicate at w comprises the n-tuples of which it is true there; and the co-extension of a predicate comprises the n-tuples of which it is false. For the present, we assume that, for all P, $\delta^+(P, w)$ and $\delta^-(P, w)$ are exclusive and exhaustive. That is:

$$\delta^+(P, w) \cap \delta^-(P, w) = \phi$$

$$\delta^+(P, w) \cup \delta^-(P, w) = D^n$$

In other words, predicates behave just as they do in standard first-order semantics, where we don't normally bother to mention co-extensions explicitly: they can just be read off from extensions. I have set things up in the way that I have for reasons that will become clear in due course.

The semantics of intentional operators is a simple generalization of the binary-relation semantics for modal operators. Thus, for any Ψ and d, R^d_Ψ is a binary relation on $\mathcal{C}$. If w and w' are in $\mathcal{C}$ then $wR^d_\Psi w'$ just if in w' things are as d (in w) Ψs them to be. Thus, for example, if Ψ is fear, then w' is a world that realizes all the fears that d (at w) has.

For particular intentional operators, Ψ, there may be reasons to put constraints on R^d_Ψ. Thus, if one thinks of Ψ as 'knows that', it is natural to require R^d_Ψ to be reflexive, so that if $t\Psi A$ is true at w, so is A. There may also be good reasons for imposing constraints that relate the various binary accessibility relations. Thus, if Φ is 'believes that', it is natural to require that if $wR^d_\Phi w'$ then $wR^d_\Psi w'$; then if $t\Psi A$ is true at w, so is $t\Phi A$ (if t knows that A then t believes that A). But I shall not pursue these details here. The reader can impose such constraints in his or her favourite way. As I have already indicated, it is the general form of a semantics that is my concern.

Truth values of sentences are assigned relative to denotations of free variables. Let s be a map from the free variables into D. Using this and δ, we can assign a denotation to every term of the language in the usual way:

if c is a constant, $\delta_s(c) = \delta\ (c)$

if x is a variable, $\delta_s(x) = s(x)$

if f is an n-place function, $\delta_s(ft_1 \ldots t_n) = \delta(f)(\delta_s(t_1), \ldots, \delta_s(t_n))$

We can now specify what it is for a sentence, A, to be true or false at world w with respect to an evaluation of the free variables, s (and an interpretation, $\mathfrak{I}$, but this will normally be taken for granted). I will write these two relations as $w \Vdash^+_s A$ and $w \Vdash^-_s A$, respectively. For atomic formulas (including propositional parameters, P):

$w \Vdash^+_s Pt_1 \ldots t_n$ iff $\langle \delta_s(t_1), \ldots \delta_s(t_n) \rangle \in \delta^+(P, w)$

$w \Vdash^-_s Pt_1 \ldots t_n$ iff $\langle \delta_s(t_1), \ldots \delta_s(t_n) \rangle \in \delta^-(P_n, w)$

For non-atomic formulas:

$w \Vdash^+_s \neg A$ iff $w \Vdash^-_s A$

$w \Vdash^-_s \neg A$ iff $w \Vdash^+_s A$

$w \Vdash^+_s A \wedge B$ iff $w \Vdash^+_s A$ and $w \Vdash^+_s B$

$w \Vdash^-_s A \wedge B$ iff $w \Vdash^-_s A$ or $w \Vdash^-_s B$

$w \Vdash_s^+ A \vee B$ iff $w \Vdash_s^+ A$ or $w \Vdash_s^+ B$
$w \Vdash_s^- A \vee B$ iff $w \Vdash_s^- A$ and $w \Vdash_s^- B$

$w \Vdash_s^+ \Box A$ iff for all $w' \in \mathcal{C}$, $w \Vdash_s^+ A$
$w \Vdash_s^- \Box A$ iff for some $w' \in \mathcal{C}$, $w \Vdash_s^- A$

$w \Vdash_s^+ \Diamond A$ iff for some $w' \in \mathcal{C}$, $w \Vdash_s^+ A$
$w \Vdash_s^- \Diamond A$ iff for all $w' \in \mathcal{C}$, $w \Vdash_s^- A$

$w \Vdash_s^+ A \rightarrow B$ iff for all $w' \in \mathcal{C}$ such that $w' \Vdash_s^+ A$, $w' \Vdash_s^+ B$
$w \Vdash_s^- A \rightarrow B$ iff for some $w' \in \mathcal{C}$, $w' \Vdash_s^+ A$ and $w' \Vdash_s^- B$

$w \Vdash_s^+ t\Psi A$ iff for all $w' \in \mathcal{C}$ such that $wR_\Psi^{\delta_s(t)}w'$, $w' \Vdash_s^+ A$
$w \Vdash_s^- t\Psi A$ iff for some $w' \in \mathcal{C}$ such that $wR_\Psi^{\delta_s(t)}w'$, $w' \Vdash_s^- A$

For reasons that I will explain in due course, I will write the quantifiers, not as $\forall$ and $\exists$, as usual, but as $\mathfrak{A}$ (for all) and $\mathfrak{S}$ (for some).

$w \Vdash_s^+ \mathfrak{S}xA$ iff for some $d \in D$, $w \Vdash_{s(x/d)}^+ A$
$w \Vdash_s^- \mathfrak{S}xA$ iff for all $d \in D$, $w \Vdash_{s(x/d)}^- A$

$w \Vdash_s^+ \mathfrak{A}xA$ iff for all $d \in D$, $w \Vdash_{s(x/d)}^+ A$
$w \Vdash_s^- \mathfrak{A}xA$ iff for some $d \in D$, $w \Vdash_{s(x/d)}^- A$

In the last four clauses, $s(x/d)$ is the evaluation of the variables that is the same as s except that its value at x is d.

The definition of logical validity is the usual one in modal logics with a base-world. If $\mathcal{S}$ is a set of sentences and A is a sentence, $\mathcal{S} \models A$ iff for every interpretation, and evaluation of the free variables, s, if $@ \Vdash_s^+ B$ for every $B \in \mathcal{S}$, $@ \Vdash_s^+ A$. $\models A$ means the same as $\phi \models A$. It is not difficult to show that if t and A contain no free variables, then $\delta_s(t)$ and $w \Vdash_s^\pm A$ do not depend on s. (For the proof of this, and the other technical claims made in this chapter, see its technical appendix. I will use $\pm$ to mean $+$ or $-$ indifferently, context sufficing to disambiguate where necessary.) In such cases, we may therefore simply drop the subscript s.

Let me make a few comments about these semantics. Given the constraints on extensions and co-extensions, we always have exactly one of $w \Vdash_s^+ A$ and $w \Vdash_s^- A$. In the present context, then, $w \Vdash_s^- A$ iff it is not the case that $w \Vdash_s^+ A$; and the truth conditions of the extensional connectives and quantifiers are just those of classical logic.

Turning to the modal elements, the modal operators, $\Box$ and $\Diamond$ are the logical operators of *S5* ($K_{\rho\tau\sigma}$). Someone who thinks that a different system of modal logic is appropriate for the logical modalities is free to modify the semantics in standard ways, e.g. by introducing an accessibility relation to be invoked in the truth/falsity conditions for the operators. But, as a matter of fact, I think that *S5* is the correct system for the logical modalities, and employing it will keep things simple. The extra complexities add nothing of substance to the topic at issue. $\rightarrow$ is the corresponding strict conditional. We can define $A \supset B$ in the usual way as $\neg A \vee B$, and so have the material conditional at our disposal too. The domain of the quantifiers is the same for every world. Thus, we have constant-domain semantics. We could have variable domain semantics, but this complicates matters and is, in any case, unnecessary. I will return to this matter in the next section. As should now be clear to anyone familiar with the elements of quantified modal logic, the logic of these semantics is constant domain *S5*.

Aside from, perhaps, the precise way I have set things up, the only real novelty is the presence of the intentional operators. Unless further constraints are put on their corresponding accessibility relations, they will each behave essentially as the modal necessity operator, $\Box$, does in constant domain *K* (where there are no constraints on the accessibility relation for $\Box$), the first argument (the agent) being carried around as a parameter. Some of the principles that the semantics validate, as can easily be checked, and which will concern us particularly, are:[12]

Logical Omniscience:
If $\models A$ then $\models t\Psi A$

Closure Under Entailment:
$t\Psi A, A \rightarrow B \models t\Psi B$

[12] For the first of these: suppose that $\not\models t\Psi A$; then there is an interpretation in which $@ \nVdash_s^+ t\Psi A$; so for some w such that $@R_\Psi^{\delta_s(t)}w$, $w \nVdash_s^+ A$. Take the interpretation that is exactly the same as this, except that its base world, $@$, is this w. In this interpretation, $@ \nVdash_s^+ A$. So $\not\models A$. For the last of these: suppose that $@ \Vdash_s^+ t\Psi\mathfrak{A}xA(x)$. Then for every w such that $@R_\Psi^{\delta_s(t)}w$, and every $d \in D$, $w \Vdash_{s(x/d)}^+ A(x)$. Consequently, for every $d \in D$, and every w such that $@R_\Psi^{\delta_s(t)}w$, $w \Vdash_{s(x/d)}^+ A(x)$. That is $@ \Vdash_s^+ \mathfrak{A}x\, t\Psi A(x)$. The other two are left as exercises.

Intentional Barcan Formula:
$\mathfrak{A}x\, t\Psi A(x) \models t\Psi\mathfrak{A}xA(x)$

Intentional Converse Barcan Formula:
$t\Psi\mathfrak{A}xA(x) \models \mathfrak{A}x\, t\Psi A(x)$

1.4 Noneism: A First Appearance

The major reason why, it is usually assumed, variable domains are more appropriate for a world semantics than constant domain, is that it seems clear that different things exist at different worlds. Thus, for example, I exist at this world, but in a world where my father was killed in the Second World War, I was never born, and so do not exist. Or conversely, and perhaps more contentiously, Sherlock Holmes does not exist at this world. But in those worlds that realize Arthur Conan Doyle's tale, he does.

But this is to assume that the denizens of a world's domain are precisely the things that exist there. And this is rejected by noneism. If one is a noneist, there would seem to be no reason why the domain of each world should not be exactly the same, namely the set of *all* objects—whatever an object's existential status at that world. This status is expressed by deploying an existence predicate. That is, we assume that there is a one-place predicate, E, such that the existent objects at a world, w, are precisely those that are in the extension of E at w, $\delta^+(E, w)$.

Of course, if one does this, one must precisely *not* read $\mathfrak{S}xA(x)$ as 'There exists something, x, such that $A(x)$'. Assuming that existence and being are the same thing, one should not even read it as 'There is something, x, such that $A(x)$'. The reading 'Something, x, is such that $A(x)$' will do nicely. This is why I have changed the symbolism: the temptation to read $\exists$ as 'there exists/is' is just too great. (There is not a similar problem with $\forall$, but I changed it to $\mathfrak{A}$ to keep $\mathfrak{S}$ company.) Thus, $\mathfrak{S}x(Px \wedge Qx)$ is: some x is such that x is a P and x is a Q. Or more simply: some Ps are Qs. One can still continue to read $\mathfrak{A}xA(x)$, as 'Every x is (or all xs are) such that $A(x)$'. Thus $\mathfrak{A}x(Px \supset Qx)$ is: every x is such that if it is a P it is a Q. Or more simply: all Ps are Qs.[13]

[13] $\supset$ is not, in fact, the right connective to be used in connection with restricted universal quantification in general, since it does not detach in inconsistent contexts. It will do here, however. $\rightarrow$ does detach, but it is too strong. The correct connective to use is neither of these, but we do not need to go into the matter here. See Beall *et al.* (Forthcoming).

If one wishes to express the more orthodox interpretation of quantifiers, one can (and has to) do this by deploying the existence predicate. Thus, 'There exists something, x, such that $A(x)$' is 'Something, x, is such that it exists and $A(x)$', i.e. $\mathfrak{S}x(Ex \wedge A(x))$. And 'Every existing thing is such that $A(x)$' is 'Everything, x, is such that if x exists then $A(x)$'. That is: $\mathfrak{A}x(Ex \supset A(x))$. I will write these as $\exists x A(x)$ and $\forall x A(x)$, respectively. Thus, these quantifiers are existentially loaded.

The admission of non-existent objects is meinongianism, or, as I shall call it—as explained in the preface to the book, and following Routley / Sylvan—*noneism*. And let me stress, as he did, that non-existent objects do not have some inferior mode of being, such as 'subsistence'. They have no mode of being whatever. They do not exist in any sense of that word (at the world in question, of course—they may, or may not, exist at others; they may even not exist at any world).

It is worth noting that even variable domain possible-world semantics appears to be committed to noneism. For the semantics themselves quantify over the objects in all the domains, not just the domain of the actual world. One can escape this conclusion by being a modal realist, and so taking every object to exist. But this gives us an extremely bloated ontology. Let us not pursue this matter at the moment; I will come back to it in a later chapter.

The noneist strategy requires us to suppose that existence is a perfectly ordinary predicate. But, tradition to the contrary not withstanding, there is no good reason not to do so. Indeed, even in classical logic, as interpreted by Quine, there is a perfectly good existence predicate: Ex is just $\exists y\ y = x$. It is just that this predicate is vacuous. Even Kant, on whom the view that existence is not a predicate is usually foisted, did not say that existence is not a predicate: he said that it is not a *determining* predicate[14]—very closely related to what meinongians have called a characterizing predicate. But if existence is a predicate, can we not run the Ontological Argument for the existence of God, and so show her to exist—and in fact, show everything to exist by an Ontological-style argument? The answer is 'no'. One needs more than just that existence is a predicate; one needs that it is a predicate that is allowed to occur in the Characterization Postulate (CP) (as that is often understood). Noneists, such as Routley, standardly reject this assumption. I will not pursue this matter now: I will return to the issue of the CP in 4.2.

[14] *Critique of Pure Reason*, A598 = B626, ff.

There are, of course, other arguments against noneism (as there are against any interesting philosophical view!). Most of them, it must be said, are very bad. And most are disposed of by Routley somewhere in the 1,000 pages of *Meinong's Jungle*. I do not intend to review all possible objections. We will return to the objections that are hardest and/or most influential in later chapters.

1.5 Worlds, Possible and Impossible

So let us leave noneism for the present, and return to the matter of worlds. As things stand so far, $Q \rightarrow Q$ is true at all worlds. Hence, $P \rightarrow (Q \rightarrow Q)$ is a logical truth. That is, given the semantics there are 'fallacies of relevance': logical truths of the form $A \rightarrow B$ where A and B share no propositional parameter. This is counter-intuitive.

The point of relevant logics is to get rid of such 'fallacies'. The major device in the world-semantics for relevant logics that delivers this is the employment of a distinctive kind of world. We distinguish within the class of closed worlds, $\mathcal{C}$, between normal and non-normal worlds (as they are often called in the literature). The normal worlds are to be thought of as (logically) possible worlds. Non-normal worlds are to be thought of as (logically) impossible worlds. The idea that there can be physically impossible worlds, that is, worlds where the laws of physics are different, is a standard one. Such worlds are still logically possible. But just as some worlds have laws of physics different from the actual physical laws, so some worlds have laws of logic different from the actual logical laws.[15] Intuitively, after all, we reason about such worlds when we consider alternative logics. Thus, a classical logician believes that the Law of Excluded Middle is valid. But they know well that if intuitionist logic were correct, this law would fail, though the law of non-contradiction would not. They therefore seem to be quite capable of considering logically impossible situations, and making discriminations about what happens within them. And given such impossible worlds, P may hold at one of them where $Q \rightarrow Q$ fails. At possible worlds, for $A \rightarrow B$ to be true,

[15] To say that $A \vDash B$ is to say, essentially: $\mathfrak{A}x$ (if x is an interpretation and @ is the base world of x, then if $@ \Vdash^+ A$, $@ \Vdash^+ B$). If what is valid can change from world to world, then the truth value of this statement must change from world to world too. Both interpretations and $\Vdash^+$ are defined set-theoretically. This change in truth value is therefore possible if the extension of the membership predicate, $\in$, may change from world to world.

we still require that at every (closed) world where A holds B holds. Hence $P \to (Q \to Q)$ is not a logical truth.

The next question is how, as a matter of technique, one arranges for $Q \to Q$ and its like to fail at a non-normal world. There are, in fact, a number of different techniques that can be deployed here. A prominent one is using a ternary relation to give the truth conditions of $\to$. Thus, we take the semantics to be furnished with a ternary relation, R, on $\mathcal{C}$; and at impossible worlds, w, we state the truth conditions of $\to$-formulas as follows:

> **RM** $w \Vdash_s^+ A \to B$ iff for all worlds $x, y \in \mathcal{C}$ such that $Rwxy$ (if $x \Vdash_s^+ A$ then $y \Vdash_s^+ B$)

('RM' is for Routley/Meyer, since this is the key move in the Routley/Meyer semantics for relevant logics.) Appropriate falsity conditions can be given, but the details need not concern us here. It is clear how we can get $Q \to Q$ to fail at a world, w. w has to be a non-normal world such that for some worlds (possible or impossible), x and y, such that $Rwxy$, Q holds at x and fails at y. With a little juggling we can, in fact, simplify matters technically. With one minimal constraint[16] we can give the truth conditions of $\to$ uniformly as RM. We will not go into details further here, though.[17]

There is another way of arranging for logical truths such as $Q \to Q$ to fail at a world, a way that is more pertinent to our present concerns. Let us turn to this. Logically impossible worlds are worlds where the laws of logic may be different. Formulas of the form $A \to B$ express entailments, laws of logic. Thus, at logically impossible worlds, one should expect such formulas to behave differently. How differently? Well, if logic can change, they can behave in pretty much any way. At an impossible world, the value of $A \to B$ might therefore be anything. Thus, in a formal model one can simply assign it an arbitrary truth value. Formulas such as $Q \to Q$ can effectively, therefore, simply be assigned the value false (and not true) at such a world.

Formally, we implement the idea as follows. An interpretation for the language is a structure $\langle \mathcal{P}, \mathcal{I}, @, D, \delta, \rangle$. $\mathcal{P}$ is the set of possible worlds, $\mathcal{I}$ the set of impossible worlds, $\mathcal{P} \cap \mathcal{I} = \phi$, $\mathcal{P} \cup \mathcal{I} = \mathcal{C}$, and $@ \in \mathcal{P}$. D and δ are exactly as before, except that at impossible worlds, δ treats formulas

[16] Namely, that if w is possible then $Rwxy$ iff $x = y$.

[17] For details, see Priest (2001), chs. 8–10.

of the form $A \rightarrow B$ essentially as atomic, assigning them extensions and co-extensions.

We have to be a little careful how we do this if quantifiers are to work properly. We need to employ a construction that will be deployed at various points in the book, so let me explain it carefully here. Suppose that the term t occurs in the formula $A(t)$. Say that t occurs *free* if it contains no occurrence of a free variable that is bound in $A(t)$. Thus, *fx* is free in *Pfx*, but not free in $\mathfrak{S}xPfx$. Call a formula a *matrix*, if all its free terms are variables, no free variable has multiple occurrences and—for the sake of definiteness—the free variables that occur in it, $x_1, \ldots, x_n$, are the least variables greater than all the variables bound in the formula, in some canonical ordering, in ascending order from left to right. Thus, for example, if P_1 is a one-place predicate, and f_2 is a two-place function symbol, then the following is a matrix provided that x_1 and x_2 are the variables that come immediately after z in a canonical ordering:

$$P_1x_1 \rightarrow \exists z P_1 f_2 z x_2$$

The important thing about matrices in the present context is that any formula can be obtained from a matrix (which might be the formula itself) by the substitution of some number of terms for the free variables, such terms being free in the result. In fact, any formula can be obtained from a unique such matrix. I will call this the matrix of the formula. If A is any formula, let $\overline{A}$ be its matrix.

Now, to return to impossible worlds, at every such world, w, δ assigns each matrix, C, of the form $A \rightarrow B$, a denotation $\delta(C, w) = \langle \delta^+(C, w), \delta^-(C, w) \rangle$; where $\delta^+(C, w), \delta^-(C, w) \subseteq D^n$ (and we continue to assume, for present, that $\delta^+(C, w)$ and $\delta^-(C, w)$ are exclusive and exhaustive).

The truth conditions for conditionals are exactly as before when $w \in \mathcal{P}$. That is:

$$w \Vdash_s^+ A \rightarrow B \text{ iff for all } w' \in \mathcal{C} \text{ such that } w' \Vdash_s^+ A,\ w' \Vdash_s^+ B$$
$$w \Vdash_s^- A \rightarrow B \text{ iff for some } w' \in \mathcal{C},\ w' \Vdash_s^+ A \text{ and } w' \Vdash_s^- B$$

(Note that the world-quantifiers still range over all the worlds in $\mathcal{C}$.) But if $w \in \mathcal{I}$, conditional formulas are treated essentially as atomic there. Thus, let $C(x_1, \ldots, x_n)$ be any matrix of the form $A \rightarrow B$, and let $t_1, \ldots, t_n$ be terms that may be freely substituted for the respective variables. Then:

$$w \Vdash_s^+ C(t_1, \ldots, t_n) \text{ iff } \langle \delta_s(t_1), \ldots \delta_s(t_n) \rangle \in \delta^+(C, w)$$
$$w \Vdash_s^- C(t_1, \ldots, t_n) \text{ iff } \langle \delta_s(t_1), \ldots \delta_s(t_n) \rangle \in \delta^-(C, w)$$

We could, in fact, make the treatment of atomic formulas uniform by having δ assign extensions and co-extensions not to each predicate, P, but to the appropriate matrix, $Px_1 \ldots x_n$, giving the truth/falsity conditions for these (at all worlds) as we have just done for conditionals at impossible worlds. But I won't backtrack to do that.

As should be clear, it is now a trivial matter to arrange for a formula of the form $Q \rightarrow Q$ to fail at an impossible world. We simply assign $\overline{Q \rightarrow Q}$ the appropriate extension.

What of the other logical machinery? The truth conditions for conjunction, disjunction, and the quantifiers remain the same at all worlds. Such operators have nothing to do with expressing laws of logic. The modal operators are clearly different, since their behaviour does concern the laws of logic, and what is logically possible or necessary at an impossible world may vary from what is actually so. Thus, at possible worlds, w, the truth and falsity conditions of the modal operators will be:

$$w \Vdash_s^+ \Box A \text{ iff for all } w' \in \mathcal{P}, w' \Vdash_s^+ A$$
$$w \Vdash_s^- \Box A \text{ iff for some } w' \in \mathcal{P}, w' \Vdash_s^- A$$

$$w \Vdash_s^+ \Diamond A \text{ iff for some } w' \in \mathcal{P}, w' \Vdash_s^+ A$$
$$w \Vdash_s^- \Diamond A \text{ iff for all } w' \in \mathcal{P}, w' \Vdash_s^- A$$

Note that the world quantifiers, as one would expect, range over only the possible worlds, $\mathcal{P}$. At impossible worlds modal formulas are treated in the same way as conditionals. Thus, if $w \in \mathcal{I}$, δ must assign each matrix of the form $\Box A$ and $\Diamond A$ an extension and co-extension at w. Then, as for formulas of the form $A \rightarrow B$, if $C(x_1, \ldots, x_n)$ is any matrix of the form $\Box A$ or $\Diamond A$, and $t_1, \ldots, t_n$ are terms that may be freely substituted for the respective variables, then:

$$w \Vdash_s^+ C(t_1, \ldots, t_n) \text{ iff } \langle \delta_s(t_1), \ldots \delta_s(t_n) \rangle \in \delta^+(C, w)$$
$$w \Vdash_s^- C(t_1, \ldots, t_n) \text{ iff } \langle \delta_s(t_1), \ldots \delta_s(t_n) \rangle \in \delta^-(C, w)$$

The definition of validity, note, remains the same, namely, truth preservation at the base world, @, in all interpretations.

One may show that these semantics have the properties required to make quantifiers work properly, and specifically that Universal Instantiation and Particular Generalization hold.

1.6 Negation

The observant reader will have noticed that I have said nothing yet about how negation behaves, now that we have introduced impossible worlds. Let me redress this lacuna.

Consider the principles of excluded middle and non-contradiction, in the form: $A \vee \neg A$ always holds; $A \wedge \neg A$ never holds. The principle of excluded middle must fail in some worlds. For either it is a not a logical truth, in which case, both A and $\neg A$ fail at some possible world; or it is a logical truth; in this case, since logic may be different at logically impossible worlds, A and $\neg A$ may both fail at some logically impossible world. In either case, the principle fails somewhere. Similarly for the principle of non-contradiction. Either A and $\neg A$ may, as a matter of logic, hold together, in which case A and $\neg A$ hold at some possible world; or they may not, as a matter of logic, do so; and since logic may be different at logically impossible worlds, A and $\neg A$ may both hold at some logically impossible world. In either case, both may hold somewhere. Allowing for the failure of these principles formally is entirely straightforward. We simply relax the constraints that $\delta^+(P, w) \cup \delta^-(P, w) = D^n$ and $\delta^+(P, w) \cap \delta^-(P, w) = \phi$—and similarly for the matrices of those formulas treated as atomic at impossible worlds. The former allows formulas of the form $A \vee \neg A$ to fail at w; the latter allows formulas of the form $A \wedge \neg A$ to hold at w.

Where we relax these constraints is another matter. The most conservative approach is to relax them only at impossible worlds. This approach delivers the semantics of a fully relevant logic. Thus, we will have $\not\models P \rightarrow (Q \vee \neg Q)$ and $\not\models (P \wedge \neg P) \rightarrow Q$. More generally, whenever $\models A \rightarrow B$ then A and B will share a predicate or propositional parameter.

The most liberal (liberating!) approach is to drop the constraint at all worlds, allowing for the actual world to contain truth value gaps and gluts. This makes the logic not only relevant but paraconsistent. That is $A, \neg A \not\models B$. For dual reasons, we also have $A \not\models B \vee \neg B$.

There is also a half-way house: relax the constraint at possible worlds in general, but retain it at the actual world, @. The actual world may, after all, be a special case. We then still have a relevant logic, but not a paraconsistent one. In this case, also, though $A \land \neg A$ holds at (the base world of) no interpretation, $\Diamond(A \land \neg A)$ may. Similarly, even though $\models A \lor \neg A$, we will not have $\models \Box(A \lor \neg A)$. Hence, the modal law of necessitation will fail.

Of course, there is no a priori reason why both constraints should be treated in the same way. Thus, those inclined to there being truth value gaps, but not to paraconsistency, will drop exhaustivity for @, but not exclusivity; those inclined to paraconsistency, but not excluded middle,[18] will do the opposite.

How one *ought* to proceed is, of course, another matter. This will turn on debates concerning whether there are truth value gaps, on dialetheism, and so on. We need not pursue these matters here. For most of this book, until the last chapter anyway, the decision makes very little difference to the matters at hand.

1.7 Open Worlds

We can now return to the issue of intentional operators, and specifically to a well-known problem. Suppose, for a moment, that we had left the semantics as they stood in 1.3, before we introduced impossible worlds. These semantics are subject to the well-known problem of *logical omniscience*. As we noted in that section, if A is a logical or necessary truth, then so is $t\Psi A$. This is just not acceptable for an arbitrary Ψ. Thus, let us suppose that the Principle of Excluded Middle is a logical truth (and if it is not, just change the example). Brouwer certainly did not believe it. Or let A be: if cows are black, cows are black. This is a logical truth, but Frege did not fear that A. Or let A be: there is an infinite number of prime numbers. 'Try to prove that' is an intentional operator; and it is certainly not true that Atilla the Hun tried to prove that A. This is the problem of logical omniscience.

It is not only logical omniscience that gives problems for intentionality. Closure under entailment does so too. This seems, clearly, equally wrong for many—if not most—intentional Ψs. Let A be $P \lor \neg P$; let B be some complex logical truth that A entails, but that no one has ever considered. I believe that A, but I don't believe that B. Or consider: the Peano postulates

[18] As in Priest (1987).

entail Fermat's Last Theorem (let us suppose). I have certainly verified that the Peano postulates hold. I have never verified that Fermat's Last Theorem holds. A final example: I may desire to eat my cake. If I eat my cake, it follows that it will no longer exist. But I do not desire that my cake no longer exist. I want to have my cake and eat it too. Irrational? Maybe, but people are like that.

Finally, though not so commonly noted, the Barcan formula and converse Barcan formula are also wrong for arbitrary intentional operators. Thus, for example, a person may know or believe of each object that it is P, but not believe that all objects are. They may not know that those are all the objects. Conversely, I may fear that nobody loves me (i.e. that $\mathfrak{A}x$ (x does not love me). But it does not follow that $\mathfrak{A}x$(I fear that x does not love me). I may not give a damn about Attila the Hun. Indeed, it may be the case that for each x I do not fear that x does not love me: what I fear is that everyone is like that.

Distinguishing impossible worlds goes some way towards resolving these problems. For let A be any logical truth, say $B \to B$. We can construct an interpretation (plus an evaluation of the free variables, s), where there is an impossible world, w, such that A fails at w (under s).[19] In that interpretation, let $@R_{\Psi}^{\delta(t)}w$. Then $@ \nVdash_s^+ t\Psi A$; so logical omniscience fails.

Unfortunately, the construction does not solve the problem of closure under entailment. For suppose that $\models A \to B$, and that in some interpretation $@ \Vdash_s^+ t\Psi A$. Then for all $w \in \mathcal{C}$ such that $@R_{\Psi}^{\delta(t)}w$, $w \Vdash_s^+ A$. But then, for all such $w \in \mathcal{C}$, $w \Vdash_s^+ B$; that is, $@ \Vdash_s^+ t\Psi B$. Nor does it solve the problems of the Barcan and converse Barcan formulas for intentional operators. For the construction modifies only the behaviour of $\to$ and the modal operators, and the Intentional Barcan formulas do not involve these at all.

What to do about this? A natural answer is as follows. Just as there are worlds that realize the way that things are conceived to be when that conception is logically possible, and worlds that realize how things are conceived to be when that conception is logically impossible, so there must be worlds that realize how things are conceived to be for the contents of arbitrary intentional states. Since such states are not closed under entailment, neither are these worlds. We are therefore led to posit a class

[19] I will not prove this here, though, since events in the rest of the section will overtake matters.

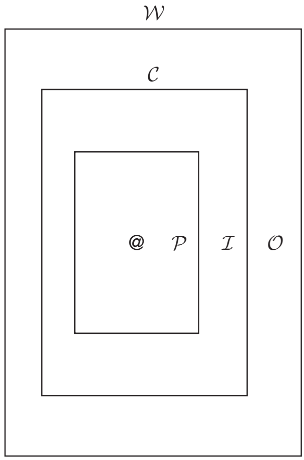

Figure 1.1

of unclosed, or *open*, worlds, $\mathcal{O}$. So let the totality of worlds, $\mathcal{W}$, will now look like Fig. 1.1.

Given how impossible worlds function, it is easy to see how open worlds should function. Given an arbitrary intentional state, there is, in general, no connection between A and B being in it, for distinct A and B. Thus, just as conditionals may behave arbitrarily at impossible worlds, all formulas may behave arbitrarily at open worlds. We may represent this formally as follows.[20]

An interpretation is now a set $\langle \mathcal{P}, \mathcal{I}, \mathcal{O}, @, D, \delta \rangle$. $\mathcal{P}$, $\mathcal{I}$, $@$ and D are as before. $\mathcal{O}$ is the set of open worlds, so $\mathcal{O} \cap \mathcal{C} = \phi$. $\mathcal{W} = \mathcal{C} \cup \mathcal{O}$. δ is as before, except that, in addition, if $w \in \mathcal{O}$, and C is *any* matrix (not just of the form $A \to B$, $\Box A$, or $\Diamond A$) then δ assigns C an extension and co-extension at w, $\delta^+(C, w)$, $\delta^-(C, w) \subseteq D^n$. (Note that we make no assumption about exclusivity and exhaustiveness—nor, clearly, should we.) At open worlds, truth/falsity conditions are given uniformly as follows. Let $C(x_1, \ldots, x_n)$ be any matrix, and $t_1, \ldots, t_n$ terms that are freely substitutable for the respective variables. Then:

$$w \Vdash_s^+ C(t_1, \ldots, t_n) \text{ iff } \langle \delta_s(t_1), \ldots, \delta_s(t_n) \rangle \in \delta^+(C, w)$$
$$w \Vdash_s^- C(t_1, \ldots, t_n) \text{ iff } \langle \delta_s(t_1), \ldots, \delta_s(t_n) \rangle \in \delta^-(C, w)$$

[20] A propositional semantics based on a similar idea, with an eye on the problem of logical omniscience, can be found in Rantala (1982).

Truth conditions at closed worlds are the same as before, except for the intentional operators. For these, worlds may be allowed to access open worlds as well, thus:

$$w \Vdash_s^+ t\Psi A \text{ iff for all } w' \in \mathcal{W} \text{ such that } wR_\Psi^{\delta(t)}w',\ w' \Vdash_s^+ A$$
$$w \Vdash_s^- t\Psi A \text{ iff for some } w' \in \mathcal{W} \text{ such that } wR_\Psi^{\delta(t)}w',\ w' \Vdash_s^- A$$

(Note, though, that in the truth/falsity conditions for $\rightarrow$ at possible worlds, the relevant class of worlds is still $\mathcal{C}$, not $\mathcal{W}$.)

Validity is still defined in terms of truth-preservation at @.

Open worlds have no effect (at @) on formulas that do not contain intentional operators. However, they suffice to destroy closure under entailment for intentional operators. Thus, suppose that $\models A \rightarrow B$. Then at every world in $\mathcal{C}$ where A holds (under some evaluation, s) so does B. However, if B is distinct from A, there may still be an open world where A holds (under s) and B does not; and @ may access this under $R_\Psi^{\delta(t)}$, in which case, $t\Psi A$ may hold at @, but not $t\Psi B$.

Thus, for example, $\models (Pa \wedge Qb) \rightarrow Pa$. Take an interpretation where:

$$\mathcal{C} = \{@\}$$
$$\mathcal{O} = \{w\}$$
$$@R_\Psi^{\delta(t)}w \text{ (and only } w\text{)}$$
$$\delta^+(Px_1 \wedge Qx_2) = D^2, \text{ but } \delta^+(Px_1) = \phi$$

(where these formulas are matrices). Then:

$$\begin{aligned} @ \Vdash^+ t\Psi\ (Pa \wedge Qb) &\Leftrightarrow w \Vdash^+ Pa \wedge Qb \\ &\Leftrightarrow \langle\delta(a), \delta(b)\rangle \in \delta^+(Px_1 \wedge Qx_2, w) \end{aligned}$$

which is true. But:

$$\begin{aligned} @ \Vdash^+ t\Psi Pa &\Leftrightarrow w \Vdash^+ Pa \\ &\Leftrightarrow \langle\delta(a)\rangle \in \delta^+(Px_1, w) \end{aligned}$$

which is not true.

The semantics also solve the problems concerning the Barcan formulas. For example, $t\Psi\mathfrak{A}xPx \not\models \mathfrak{A}x\, t\Psi Px$. To see this, consider an interpretation where

$$D = \{a\}$$
$$\mathcal{C} = \{@\}$$
$$\mathcal{O} = \{w\}$$
$$@R_{\Psi}^{\delta_s(t)} w \text{ (and nothing else)}$$
$$\delta^+(\mathfrak{A}xPx, w) = \{\langle\rangle\}$$
$$\delta^+(Px, w) = \phi$$

(where Px is a matrix).

$$@ \Vdash_s^+ t\Psi\mathfrak{A}xPx \quad \Leftrightarrow \quad w \Vdash_s^+ \mathfrak{A}xPx$$
$$\Leftrightarrow \quad \langle\rangle \in \delta^+(\mathfrak{A}xPx, w)$$

which is true. But:

$$@ \Vdash_s^+ \mathfrak{A}x\, t\Psi Px \quad \Leftrightarrow \quad \text{for all } d \in D,\ @ \Vdash_{s(x/d)}^+ t\Psi Px$$
$$\Leftrightarrow \quad \text{for all } d \in D,\ w \Vdash_{s(x/d)}^+ Px$$
$$\Leftrightarrow \quad \text{for all } d \in D,\ d \in \delta^+(Px, w)$$

which is not true. For a counter-model for the converse implication, simply set $\delta^+(\mathfrak{A}xPx, w) = \phi$, and $\delta^+(Px, w) = D$. Thus, all the problems we have noted are solved.

One might well worry that this solution is too cheap. It destroys all inferences concerning intentional operators. Well, not quite all. We still have various inferences concerning quantifiers, e.g.: $t\Psi Pc \models \mathfrak{S}xt\Psi Px$. We also have those inferences that hold in virtue of any constraints we impose on the accessibility relation. Thus, if R_{Ψ}^d is reflexive (for all d), then $t\Psi A \models A$, etc. And maybe, as I explained in 1.3, for some intentional notions one should demand more inferences. Thus, if Ψ is 'knows that', it is natural (though I think mistaken) to suppose that it, at least, is closed under entailment. If this is right, it can be accommodated by placing appropriate constraints on R_{Ψ}^d, namely that it accesses only worlds that are closed under entailment (e.g. worlds in $\mathcal{C}$). There certainly are intentional notions that are closed under some notion of logical consequence. Thus, 'x is rationally committed to it being the case that' is closed under whatever logic x accepts. One logically closed intentional notion will play an important role in Ch. 4. But examples of this kind are special cases;

generally speaking, intentional notions *are* pretty anarchic logically. That is just the nature of the beasts.[21]

Finally, one might note, open worlds serve to invalidate logical omniscience just as much as closure under entailment and the Barcan formulas. Thus, if A is a logical truth, it is trivial to construct an open world where A does not hold. Indeed, as the semantics makes clear, impossible worlds can be thought of as open worlds of a certain kind (where only the conditionals and modals are anarchic). Why not simply dispense with impossible worlds altogether? As far as the intentional operators goes, one can, in fact, do this. The non-intentional logic is then constant-domain $S5$, rather than some constant domain relevant logic. But, arguably, relevant logic is a much better account of entailment than $S5$.[22] So one ought to countenance such worlds anyway. And the above construction shows, at the very least, that relevant logic is quite compatible with the appropriate behaviour of intentional operators.

1.8 Conclusion

In this chapter we have made a start on intentionality. We have looked at a semantics for intentional operators that deploys worlds of various kinds. These not only accommodate relevant logic, but also are not subject to the problems of logical omniscience and closure under entailment. I have also gestured in the direction of noneism. The problems concerning intentionality have only just started, however. In this chapter we have ignored identity. Identity brings problems of its own. In the next chapter we turn to these.

1.9 Technical Appendix

The propositional logic determined by the above semantics is that called N_4 in Priest (2001), ch. 9. The logic of the intentional operators, Ψ, is almost trivial, as I have already noted. The only thing, therefore, that

[21] Some (e.g. Horcutt 1972), have wondered whether something that verifies so few inferences concerning an intentional notion, and specifically 'knows that', is worth calling a *logic* at all. Perhaps not, though even the null logic is, strictly speaking, a logic. More to the point, the fact that the logic is relatively uninteresting does not mean the semantics is uninteresting. It is, in fact, a hard matter to give an account of the semantics of intentional operators that shows why various inferences *fail*.

[22] See e.g. Routley *et al.* (1982).

requires much comment is the quantifiers. In this technical appendix I show that these work appropriately in the semantics given. Because of the slightly slippery behaviour of worlds, and because the proofs of subsequent formal appendixes will build on these proofs, I will spell them out in full detail. In many of the cases, the proofs for truth $(+)$ and falsity $(-)$ are virtually the same. Where this is so, I will write $\pm$ and leave the reader to disambiguate.

Lemma 1 *Fix any interpretation. Let t and A be any term and formula. Then if s_1 and s_2 are any evaluations of the variables that agree on the variables free in t and A:*

1. $\delta_{s_1}(t) = \delta_{s_2}(t)$
2. for all $w \in \mathcal{W}$, $w \Vdash^{\pm}_{s_1} A \Leftrightarrow w \Vdash^{\pm}_{s_2} A$

Proof The proof of 1 is by recursion on the way that terms are constructed. For constants:

$$\delta_{s_1}(c) = \delta\ (c) = \delta_{s_2}(c)$$

For variables:

$$\delta_{s_1}(x) = s(x) = \delta_{s_2}(x)$$

For function symbols:

$$\begin{aligned} \delta_{s_1}(ft_1 \ldots t_n) &= \delta(f)(\delta_{s_1}(t_1), \ldots, \delta_{s_1}(t_n)) \\ &= \delta(f)(\delta_{s_2}(t_1), \ldots, \delta_{s_2}(t_n)) \quad \text{by induction hypothesis (IH)} \\ &= \delta_{s_2}(ft_1 \ldots t_n) \end{aligned}$$

The proof of 2 is in two cases, depending on whether w is in $\mathcal{C}$ or $\mathcal{O}$. Suppose that $w \in \mathcal{C}$. The argument then is by recursion, using 1 for the basis case:

$$\begin{aligned} w \Vdash^{\pm}_{s_1} Pt_1 \ldots t_n &\Leftrightarrow \langle \delta_{s_1}(t_1), \ldots, \delta_{s_1}(t_n) \rangle \in \delta^{\pm}(P, w) \\ &\Leftrightarrow \langle \delta_{s_2}(t_1), \ldots, \delta_{s_2}(t_n) \rangle \in \delta^{\pm}(P, w) \quad \text{by 1} \\ &\Leftrightarrow w \Vdash^{\pm}_{s_2} Pt_1 \ldots t_n \end{aligned}$$

$$\begin{aligned} w \Vdash^{\pm}_{s_1} \neg A &\Leftrightarrow w \Vdash^{\mp}_{s_1} A \\ &\Leftrightarrow w \Vdash^{\mp}_{s_2} A \quad \text{by IH} \\ &\Leftrightarrow w \Vdash^{\pm}_{s_2} \neg A \end{aligned}$$

The cases for $\vee$ and $\wedge$ are similar.

For $\rightarrow$ and the modal operators, there are two cases, depending on whether w is in $\mathcal{P}$ or $\mathcal{I}$. If $w \in \mathcal{P}$. Then:

$$\begin{array}{lll} w \Vdash^{+}_{s_1} A \rightarrow B & \Leftrightarrow & \text{for all } w' \in \mathcal{C}, \text{ if } w' \Vdash^{+}_{s_1} A \text{ then } w' \Vdash^{+}_{s_1} B \\ & \Leftrightarrow & \text{for all } w' \in \mathcal{C}, \text{ if } w' \Vdash^{+}_{s_2} A \text{ then } w' \Vdash^{+}_{s_2} B \quad \text{by IH} \\ & \Leftrightarrow & w \Vdash^{+}_{s_2} A \rightarrow B \end{array}$$

$$\begin{array}{lll} w \Vdash^{-}_{s_1} A \rightarrow B & \Leftrightarrow & \text{for some } w' \in \mathcal{C}, w' \Vdash^{+}_{s_1} A \text{ and } w' \Vdash^{-}_{s_1} B \\ & \Leftrightarrow & \text{for some } w' \in \mathcal{C}, w' \Vdash^{+}_{s_2} A \text{ and } w' \Vdash^{-}_{s_2} B \quad \text{by IH} \\ & \Leftrightarrow & w \Vdash^{+}_{s_2} A \rightarrow B \end{array}$$

For $\Box$:

$$\begin{array}{lll} w \Vdash^{\pm}_{s_1} \Box A & \Leftrightarrow & \text{for all/some } w' \in \mathcal{P}, w' \Vdash^{\pm}_{s_1} A \\ & \Leftrightarrow & \text{for all/some } w' \in \mathcal{P}, w' \Vdash^{\pm}_{s_2} A \quad \text{by IH} \\ & \Leftrightarrow & w \Vdash^{\pm}_{s_2} \Box A \end{array}$$

The case for $\Diamond$ is similar.

Now suppose that $w \in \mathcal{I}$. Consider a formula of the form $C(t_1, \ldots, t_n)$, where C is a matrix of the form $A \rightarrow B$, $\Box A$, or $\Diamond A$; each t_i may contain some of $x_1, \ldots, x_n$ free, and is free when substituted in C. Then:

$$\begin{array}{lll} w \Vdash^{\pm}_{s_1} C(t_1, \ldots, t_n) & \Leftrightarrow & \langle \delta_{s_1}(t_1), \ldots, \delta_{s_1}(t_n) \rangle \in \delta^{\pm}(C, w) \\ & \Leftrightarrow & \langle \delta_{s_2}(t_1), \ldots, \delta_{s_2}(t_n) \rangle \in \delta^{\pm}(C, w) \quad \text{by 1} \\ & \Leftrightarrow & w \Vdash^{\pm}_{s_2} C(t_1, \ldots, t_n) \end{array}$$

For the intentional operators, $w \Vdash^{\pm}_{s_1} t\Psi A$:

$$\begin{array}{ll} \Leftrightarrow & \text{for all/some } w' \in \mathcal{W} \text{ such that } wR_{\Psi}^{\delta_{s_1}(t)} w', w' \Vdash^{\pm}_{s_1} A \\ \Leftrightarrow & \text{for all/some } w' \in \mathcal{W} \text{ such that } wR_{\Psi}^{\delta_{s_2}(t)} w', w' \Vdash^{\pm}_{s_2} A \\ & \qquad\qquad\qquad\qquad\qquad\qquad \text{by 1 and IH} \\ \Leftrightarrow & w \Vdash^{\pm}_{s_2} t\Psi A \end{array}$$

And for the quantifiers:

$$\begin{array}{lll} w \Vdash^{\pm}_{s_1} \mathfrak{A}yA(y) & \Leftrightarrow & \text{for all/some } d \in D, w \Vdash^{\pm}_{s_1(y/d)} A(y) \\ & \Leftrightarrow & \text{for all/some } d \in D, w \Vdash^{\pm}_{s_2(y/d)} A(y) \quad \text{by IH} \\ & \Leftrightarrow & w \Vdash^{\pm}_{s_1} \mathfrak{A}yA(y) \end{array}$$

The case of $\mathfrak{S}$ is similar.

The second case is where $w \in \mathcal{O}$. In this case, there is no recursion at all, and the argument for each formula is the same as that for $\rightarrow$ and modal operators at impossible worlds. ■

Lemma 2 *Fix any interpretation. Let $t'(x)$ and $A(x)$ be any term and formula. Let t be any term that can be freely substituted for x in these. Let s be any evaluation of the free variables, then if $d = \delta_s(t)$:*

1. $\delta_{s(x/d)}(t'(x)) = \delta_s(t'(t))$
2. for all $w \in \mathcal{W}$, $w \Vdash^{\pm}_{s(x/d)} A(x) \Leftrightarrow w \Vdash^{\pm}_{s} A(t)$

Proof The proof of 1 is by recursion on the way that terms are constructed. If $t'(x)$ is either a constant, or a variable distinct from x, the result follows by Lemma 1. For x:

$$\delta_{s(x/d)}(x) = d = \delta_s(t)$$

For function symbols:

$$\begin{aligned} \delta_{s(x/d)}(ft_1(x)\ldots t_n(x)) &= \delta(f)(\delta_{s(x/d)}(t_1(x)),\ldots,\\ &\qquad \delta_{s(x/d)}(t_n(x))\,)\\ &= \delta(f)(\delta_s(t_1(t)),\ldots,\delta_s(t_n(t))\,) \qquad \text{by IH}\\ &= \delta_s(ft_1(t)\ldots t_n(t)) \end{aligned}$$

The proof of 2 is in two cases, depending on whether w is in $\mathcal{C}$ or $\mathcal{O}$. Suppose that $w \in \mathcal{C}$. The argument then is by recursion, using 1 for the basis case.

$$\begin{aligned} & w \Vdash^{\pm}_{s(x/d)} Pt_1(x)\ldots t_n(x):\\ \Leftrightarrow\;& \langle \delta_{s(x/d)}(t_1(x)),\ldots,\delta_{s(x/d)}(t_n(x))\rangle \in \delta^{\pm}(P,w)\\ \Leftrightarrow\;& \langle \delta_s(t_1(t)),\ldots,\delta_s(t_n(t))\rangle \in \delta^{\pm}(P,w) \qquad \text{by 1}\\ \Leftrightarrow\;& w \Vdash^{\pm}_{s} Pt_1(t)\ldots t_n(t) \end{aligned}$$

For $\neg$:

$$\begin{aligned} w \Vdash^{\pm}_{s(x/d)} \neg A(x) &\Leftrightarrow w \Vdash^{\mp}_{s(x/d)} A(x)\\ &\Leftrightarrow w \Vdash^{\mp}_{s} A(t) \qquad \text{by IH}\\ &\Leftrightarrow w \Vdash^{\pm}_{s} \neg A(t) \end{aligned}$$

The cases for $\vee$ and $\wedge$ are similar.

For $\rightarrow$ and the modal operators, there are two cases, depending on whether w is in $\mathcal{C}$ or $\mathcal{O}$. Suppose that $w \in \mathcal{C}$:

$$\begin{aligned} w \Vdash^{+}_{s(x/d)} A(x) \rightarrow B(x) &\Leftrightarrow \text{for all } w' \in \mathcal{C}, \text{if } w' \Vdash^{+}_{s(x/d)} A(x)\\ &\qquad \text{then } w' \Vdash^{+}_{s(x/d)} B(x)\\ &\Leftrightarrow \text{for all } w' \in \mathcal{C}, \text{if } w' \Vdash^{+}_{s} A(t)\\ &\qquad \text{then } w' \Vdash^{+}_{s} B(t) \qquad \text{by IH}\\ &\Leftrightarrow w \Vdash^{+}_{s} A(t) \rightarrow B(t) \end{aligned}$$

The case for falsity is similar.

$$\begin{array}{rcll} w \Vdash^{\pm}_{s(x/d)} \Box A(x) & \Leftrightarrow & \text{for all/some } w' \in \mathcal{C},\ w' \Vdash^{\pm}_{s(x/d)} A(x) & \\ & \Leftrightarrow & \text{for all/some } w' \in \mathcal{C},\ w' \Vdash^{\pm}_{s} A(t) & \text{by IH} \\ & \Leftrightarrow & w \Vdash^{\pm}_{s} \Box A(t) & \end{array}$$

The cases for $\Diamond$ are similar.

Now suppose that $w \in \mathcal{I}$. Every formula of the form $A \to B$, $\Box A$ and $\Diamond A$ is of the form $C(t_1, \ldots, t_n)$, where C is a matrix, each t_i may contain x free and is free when substituted in C. Then $w \Vdash^{\pm}_{s(x/d)} C(t_1(x), \ldots, t_n(x))$:

$$\begin{array}{cll} \Leftrightarrow & \langle \delta_{s(x/d)}(t_1(x)), \ldots, \delta_{s(x/d)}(t_n(x)) \rangle \in \delta^{\pm}(C, w) & \\ \Leftrightarrow & \langle \delta_s(t_1(t)), \ldots, \delta_s(t_n(t)) \rangle \in \delta^{\pm}(C, w) & \text{by 1} \\ \Leftrightarrow & w \Vdash^{\pm}_{s} C(t_1(t), \ldots, t_n(t)) & \end{array}$$

For the intentional operators, $w \Vdash^{\pm}_{s(x/d)} t'(x)\Psi A(x)$:

$$\begin{array}{cll} \Leftrightarrow & \text{for all/some } w' \in \mathcal{W} \text{ such that } wR_{\Psi}^{\delta_{s(x/d)}(t'(x))}w', & \\ & w' \Vdash^{\pm}_{s(x/d)} A(x) & \\ \Leftrightarrow & \text{for all/some } w' \in \mathcal{W} \text{ such that } wR_{\Psi}^{\delta_{s}(t'(t))}w', & \\ & w' \Vdash^{\pm}_{s} A(t) & \text{by 1 and IH} \\ \Leftrightarrow & w \Vdash^{\pm}_{s} t'(t)\Psi A(t) & \end{array}$$

For the universal quantifier, suppose that $A(x)$ is of the form $\mathfrak{A}yB(x)$. If y is x then $\mathfrak{A}yB(t)$ is just $\mathfrak{A}yB(x)$, and x is not free in this formula, so the result follows from Lemma 1. So suppose that x and y are distinct. Then:

$$\begin{array}{rcll} w \Vdash^{\pm}_{s(x/d)} \mathfrak{A}yB(x) & \Leftrightarrow & \text{for all/some } e \in D,\ w \Vdash^{\pm}_{s(x/d,y/e)} B(x) & \\ & \Leftrightarrow & \text{for all/some } e \in D,\ w \Vdash^{\pm}_{s(y/e)} B(t) & (\star) \\ & \Leftrightarrow & w \Vdash^{\pm}_{s} \mathfrak{A}yB(t) & \end{array}$$

For $(\star)$, since t is free when substituted for x, it cannot contain y free. Hence, $\delta_s(t) = \delta_{s(y/e)}(t)$ by Lemma 1, and we can apply the induction hypothesis where s is $s(y/e)$.

The case for $\mathfrak{S}$ is similar.

The other case is when $w \in \mathcal{O}$. In this case, there is no recursion at all, and the argument for each formula is the same as that for $\to$ and modal operators at impossible worlds. ■

Lemma 2 is exactly what one needs to show that quantifiers behave properly. In particular:

Corollary 3 *If t is free when substituted for x in* $A(x)$ *then:*

1. $\mathfrak{A}xA(x) \models A(t)$
2. $A(t) \models \mathfrak{S}xA(x)$

Proof For 1, suppose that $@ \Vdash_s^+ \mathfrak{A}xA(x)$. Then for all $d \in D$, $@ \Vdash_{s(x/d)}^+ A(x)$. Let $d = \delta_s(t)$. Then $@ \Vdash_s^+ A(t)$ by Lemma 2. For 2, suppose that $@ \Vdash_s^+ A(t)$. Then where $d = \delta_s(t)$, $@ \Vdash_{s(x/d)}^+ A(x)$ by Lemma 2. Hence, $@ \Vdash_s^+ \mathfrak{S}xA(x)$. ■

2
Identity

2.1 Introduction: Identity and Intentionality

The previous chapter looked at and resolved a number of the problems that arise in connection with intentional operators. There is one further problem that needs to be addressed, however. This concerns the behaviour of identity with respect to such operators.

We will approach the issue via a paradox traditionally associated with the Ancient Megarian logician, Eubulides. After explaining the paradox, we will look at various aspects of it, and I will give a semantics for identity that resolves the problem. First, however, let us note how one would most naturally add identity to the semantics of the previous chapter.

2.2 Adding Identity

To do this, we add a binary predicate, $=$, to the language, and give its semantics as follows. δ assigns $=$ an extension and co-extension. At possible worlds, $w \in \mathcal{P}$:

$$\delta^+(=, w) = \{\langle d, d \rangle : d \in D\}$$

Note that this is world-invariant. The co-extension of $=$ is also a world-invariant set; and, we may assume (at least for the present), $\delta^+(=, w)$ and $\delta^-(=, w)$ are exclusive and exhaustive. These, then are the classical conditions for identity. Thus, at possible worlds, identity behaves in an orthodox fashion. In particular, $\models t = t$.

At impossible worlds, laws of logic may fail. In particular, instances of the law of identity may fail. To achieve this for $\rightarrow$-formulas, we allowed them to behave arbitrarily. We may do the same for identity. Thus, if $w \in \mathcal{I}$, all we require is that $\delta^+(=, w) \subseteq D^2$ and $\delta^-(=, w) \subseteq D^2$. There is no requirement that the extension and co-extension be world-invariant, that identity statements behave consistently, or that instances of the law of identity hold at an impossible world. This makes identity an upstanding denizen of a good relevant logic. Thus, for example, we do not have $\models (t_1 = t_2 \wedge \neg t_1 = t_2) \rightarrow A$ or $\models A \rightarrow t = t$.

Finally, at open worlds, formulas containing the identity predicate work as in 1.7: they are treated as atomic formulas. Thus, any entailment consequences concerning sentences containing identity are broken at these worlds, as required for the failure of the closure of intentional states.

Even given all these things, the semantics still suffice to verify not only the law of identity, but the substitutivity of identicals, or SI as we will call it:

$$t_1 = t_2, A(t_1) \models A(t_2)$$

provided that t_1 and t_2 are free when substituted in $A(x)$ (see the appendix to this chapter). In this regard, all the hard work is done by the extension of $=$ at $@$. And it is SI that will concern us in the rest of the chapter: it is the core of paradox, as we will now see.

2.3 Eubulides the Paradoxer

The most famous paradoxer of Antiquity is undoubtedly Zeno. His paradoxes, particularly those of motion, have exercised philosophers since he formulated them. But, to my mind, the greatest paradoxer of Antiquity was not Zeno but the Megarian philosopher Eubulides. Eubulides is reputed to have formulated seven paradoxes, which Diogenes Laertius lists as: the Liar, the Disguised, the Electra, the Veiled Figure, the Sorites, the Horned One, and the Bald Head.[1] It would appear that some of these were variants of the others, and that there were basically four different

[1] Hicks (1925), ii. 108. Naturally, one can dispute whether Eubilides really did invent these paradoxes. For example, some have attributed the Disguised (the Hooded Man) to Euclides, the founder of the Megarian school.

paradoxes, which are:[2]

1. *The Liar*. 'A man says that he is lying. Is what he says true or false?'
2. *The Hooded Man, the Unnoticed Man, the Electra*. 'You say you know your brother. But that man who came in just now with his head covered is your brother, and you do not know him.'
3. *The Bald Man, or the Heap*. 'Would you say that a man was bald if he had only two hairs? Yes. Would you . . . , etc. Then where do you draw the line?'
4. *The Horned Man*. 'What you have not lost you still have. But you have not lost horns. So you still have horns.'

Eubulides' arguments must have seemed like sophisms to many of his contemporaries, and made him an easy target for parody. Indeed, a contemporary Comic poet wrote:[3] 'Eubulides the Eristic, who propounded his quibbles about horns and confounded the orators with falsely pretentious arguments, is gone with all the braggadocio of a Demosthenes.' But from the perspective of two and a half thousand years later, this low opinion is hardly justified.

The fourth of the above paradoxes is certainly little more than a sophism. It employs a device that is often used by barristers and other tricksters, and would now be classified as a *Fallacy of Many Questions*, of the kind 'Have you stopped beating your wife?' Literally, if you never had horns then you never lost them. Thus, the conditional 'If you have not lost horns you (still) have them' is false. The trick gets its bite from the conversational implicatures generated by the sorts of context in which one would normally talk of loss. The first and third paradoxes, the Liar and the Sorites are, by contrast, quite different. As no one familiar with contemporary philosophical logic needs to be told, these are of central importance to contemporary debates. Moreover, two and a half thousand years since Eubulides, there is still no consensus at all as to how to solve either of these paradoxes. This attests to their profundity. Compare the situation with that concerning Zeno's paradoxes. Though philosophers may still argue about them, there has been, for at least a century, a general consensus concerning the solution to these paradoxes. This is why I said that, of Zeno and Eubulides, it is the latter who is the greater.

[2] See Kneale and Kneale (1962), 114, who cite the classical sources.

[3] Hicks (1925), ii. 108.

2.4 The Hooded Man Paradox

What of Eubulides' second paradox, the Hooded Man? It does not have the notoriety of the Liar or the Sorites, but it is, none the less, one of the fundamental paradoxes of intentionality. It is this which will concern us.

Let us start with a clean formulation. We suppose that a man walks into the room. The man is wearing a hood, and unbeknownst to you, it is your brother. Then the argument is simply:

> This man is your brother.
> You do not know this man.
> ―――――――――――
> You do not know your brother.

The premises are true, but the conclusion appears to be untrue. Yet the argument is an instance of the SI.

There is an easy solution here, though. The second premise is false. You *do* know this man. You just don't realize that you do. More of this in the next chapter. But this solution is too swift. Though you may, in fact, know the man, you don't know who he is, at least while he wears the hood. But you do know who your brother is. So we have the following:

> This man is your brother.
> You do not know who this man is.
> ―――――――――――――
> You do not know who your brother is.

This is another instance of SI, though the premises seem true and the conclusion false.

The argument raises many issues. Let's see if we can get rid of some of the less central ones. For a start, what is it to know who somebody is? Suppose that you ask a child who Jack the Ripper was. They say 'He was a person in Victorian London who was notorious for murdering and disembowelling prostitutes, though his identity was never discovered.' The child knows who the Ripper was. But suppose that you ask a historian the same question. They know all that the child knows, but they will tell you truly that we do not know who the Ripper was. To do that, we would have to know something like: the Ripper was Queen Victoria, or the Ripper was Conan Doyle.

What we need to know to know who someone is, is, therefore, context-dependent. But whatever the context, knowing who someone is comes down to knowing certain things about them, that is, knowing *that* they are so and so, and such and such. In the case of Eubulides' paradox, knowing who the hooded man was would be things like knowing that his name was such and such, knowing that without the hood he looked like so and so, and knowing, indeed, that he was your brother.[4] The paradox can therefore be reformulated in terms of *knowing that*. Take any one of these identifying properties—say, for the sake of example, the property of being born in Megara. Then the paradoxical argument comes down to the following and its like:

This man is your brother.
You do not know that this man was born in Megara.
___
You do not know that your brother was born in Megara.

Next: Eubulides' argument employs demonstratives, and especially the demonstrative 'this man'. The denotation of a demonstrative depends on context. In this case, the referent of the demonstrative is fixed by the intention of the utterer. But if the context does not change, the denotation of a demonstrative does not change, and the same role can be played by a name referring to the object in question. And the context does not change in this argument, so we can simply ignore the extra complexity created by the demonstrative, and take the person to be referred to by a name. Let us, therefore, christen the hooded man 'Nescio', where this is a rigid designator. (That is, a term whose denotation remains the same in all worlds.) The argument then becomes:

Nescio is your brother.
You do not know that Nescio was born in Megara.
___
You do not know that your brother was born in Megara.

[4] Hintikka (1962), 132, claims that 'you know who *a* is' is to be understood as: $\exists x K x = a$ (where K is 'You know that'). This is dubious. Not only does it not make *knowing who* context-independent, but it requires the failure of Existential Generalization, even if *a* is a rigid designator, since, presumably, it is always true that $Ka = a$. But even this understanding cashes out *knowing who* in terms of *knowing that*. The most careful analysis of *knowing who* of which I am aware is provided by Boër and Lycan (1986). They recognize the context-dependence of *knowing who*, taking the relevant contexts to be certain speaker-purposes. They argue that *knowing who* is a certain kind of *knowing that*, and provide a sensitive analysis of the kind of *knowing that* that it is. The details need not concern us here. It should be noted that their analysis is greatly complicated by their employment of a paratactic analysis of *knowing that*, rather than the much more straightforward one employed in this book.

Or to contrapose, simplifying again:

Nescio is your brother.
You know that your brother was born in Megara.
———
You know that Nescio was born in Megara.

In what follows, when I refer to the Hooded Man argument, it is this argument to which I will be referring.

And with this version of the argument, we cannot avoid the problem, as we did the original, by saying that you *do* know that Nescio was born in Megara; you just don't realize this. For you certainly do realize that your brother was born in Megara. Hence, the problem would then reappear with a different example of the same kind:

Nescio is your brother.
You realize that your brother was born in Megara.
———
You realize that Nescio was born in Megara.

And if you still doubt, consider the fact that the people of the thirteenth century did not know that water was H_2O. They did know that water was water. Or that one knows a priori that George Eliot was George Eliot, but one does not know a priori that George Eliot was Marian Evans.[5]

Before we move on to a solution, let us note the further, and important, fact that there is nothing specific about knowledge in this context. Paradoxes of the kind in question arise with *all* intentional operators. If this is

[5] Salmon (1986) has an account of intensional contexts that allows substitutivity universally in such contexts. According to him, knowledge—and similar intensional operators—are, *in stricto sensu*, ternary relationships between an agent, a proposition, and a guise (or Fregean sense). Knowledge *simpliciter* is always knowledge relative to *some* guise. Hence, he thinks that you do realize that Nescio was born in Megara (since you know it relative to Nescio's guise *your brother*). You do not know it relative to the guise *man who has just entered the room*. Moreover, Salmon claims, the proposition that George Eliot is George Eliot is a priori. Hence (*contra* Kripke), so is the proposition that George Eliot is Marian Evans. It seems to me that someone who accepts this has lost contact with the work that a priority needs to do. There is no way that that particular fact could be reasoned out without empirical knowledge. One might be tempted to say that there is some guise under which it could not be reasoned out without empirical knowledge; but for Salmon, a priority is a property of propositions; it is not relative to a guise (p. 133). Nor can one say that it is the truth of the sentence 'George Eliot is Marian Evans' that cannot be reasoned out without empirical knowledge: the truth of no sentence can be reasoned out without empirical knowledege (about meanings).

not clear, just consider the following arguments:

Nescio is your brother.
You believe that your brother was born in Megara.
———
You believe that Nescio was born in Megara.

The conclusion may well be false—especially if you don't know that Nescio is your brother. Or:

Nescio is your brother.
You fear that Nescio is now dead.
———
You fear that your brother is now dead.

Again, the conclusion may well be untrue. Or:

Nescio is your brother.
You desire that Nescio die.
———
You desire that your brother die.

And so on: just recall Oedipus and Jocasta.

What is at issue, then, has nothing to do with knowledge as such. It is an issue about the behaviour of identity in all intentional propositional contexts. The semantics that follow, then, will be quite general, though I shall often illustrate informally with knowledge. With this in mind, back to the argument.

2.5 Descriptions and Rigid Designators

There is still one more preliminary issue that needs to be addressed. This is how to understand the noun-phrase 'your brother'. This is a demonstrative, but it is not a simple demonstrative, since it packs in the information that the thing referred to has a certain property—that of being your brother. Since the context is not changing, we can ignore the demonstrative aspect, but the question is how this description is supposed to be functioning. A common semantics for descriptions makes them non-rigid designators. That is, their denotations may vary from world to world. It is well known that SI (along with quantifier inferences such as Particular Generalization) may fail in non-extensional contexts when the term at issue is a description of this kind.[6] For example, it is necessarily

[6] See e.g. Fitting and Mendelsohn (1998), esp. chs. 9, 12.

the case that $9 = 9$; but it is not necessarily the case that the number of planets $= 9$. Though 'the number of the planets' refers to 9 in this world, in other worlds it may refer to a different number.

If one adopts this approach to descriptions, then the paradox is solved. In Ch. 4 I will give a semantics for descriptions, and I will treat them as rigid designators. This solution is not, therefore, open here. But even if one were to adopt a theory of descriptions according to which they were not rigid, one would still not have resolved the issue. For the problems with SI arise just as much with designators that are rigid, such as proper names. Thus, suppose that your brother's name is 'Cain'. We can simply use this in formulating the argument. The argument is now:

Nescio is Cain.
You know that Cain was born in Megara.
———
You know that Nescio was born in Megara.

This is of the form:

$$\frac{\begin{array}{l} n = c \\ a\Psi Mc \end{array}}{a\Psi Mn}$$

(where a is you). And as we have already noted, this instance of SI is valid in the semantics for identity of 2.2. Yet this seems wrong. You *do* know that Cain was born in Megara. You *don't* know that Nescio was. Hence, despite the semantics, this instance of SI does seem invalid.[7]

2.6 The Puzzle about Pierre

Before we look at a solution to the problem, it is worth noting that the view that examples of the kind with which we are dealing demonstrate the failure of SI in epistemic contexts has been challenged by Kripke (1979). He argues that contradictions of the kind in question arise even without SI. It is therefore wrong to point the finger of suspicion at it. His well-known

[7] One might maintain that all names are really covert descriptions, and thus reduce this version of the argument to the one with descriptions. But such a move is well known to face grave difficulties. See Kripke (1972). It is particularly hard to suppose that demonstratives are covert descriptions, since, for these, uptake of reference may be secured with no linguistic intermediary at all.

example concerns a native French-speaker, Pierre, who expresses one of his beliefs by saying 'Londres est jolie'. He then learns English, and comes to express one of his beliefs by saying 'London is not pretty'—without revoking any former dispositions concerning his assertions in French: he simply does not realize that 'Londres' and 'London' refer to the same place. He would appear to believe that London is both pretty and not pretty. He may even vehemently reject the claim 'London is pretty', in which case he would seem both to believe and not to believe that London is pretty.

In fact, the detour through French is unnecessary, as Kripke points out. The issue would be just the same as that which arises for a person, say Pierre, who sincerely asserts: 'George Eliot is a man' and 'Marian Evans is not a man'—or, even stronger, denies 'Marian Evans is a man'—unaware that they are the same person.

Let us concentrate on the monolingual version. In the given situation, it is virtually irresistible to hold that:

> Pierre believes George Eliot to be a man

and that:

> Pierre believes Marian Evans not to be a man

—or, in the stronger case, that:

> Pierre does not believe Marian Evans to be a man

We have this from the horse's mouth; and though this sort of evidence may be overridden in some cases (e.g. when speakers do not properly understand the words they use) we can set up the situation in such a way that cases of this kind are explicitly ruled out. Contradiction looms here when, and only when, we add the further premise that Eliot is Evans, to conclude that Pierre believes Eliot not to be a man—or, in the stronger case, does not believe Eliot to be a man. SI is essentially involved here.

The contradictions that Kripke points to, by contrast, concern not this, but how Pierre's beliefs may be reported in *paraphrase*. If we paraphrase Pierre's beliefs about Evans by using the name 'Eliot' we have similar contradictions. Now we often paraphrase people's views in reporting them. Suppose that you tell me that it was the author of the Sherlock

Holmes stories who was the Ripper. It would normally be fair for me to report your belief to a third party by saying that you think that Doyle was the Ripper. If, however, you thought also that Doyle stole the Holmes stories, this would not be fair paraphrase.

Similarly, and closer to hand, suppose that it is common knowledge in a group (which includes Pierre) that Eliot is Evans, and that Pierre believes Eliot to be a man (perhaps believing that Eliot was, in fact, a very successful transvestite), then it would be quite legitimate to report his belief by saying that he believed Evans to be a man. But in the sort of case in question, where Pierre does not know that Eliot and Evans are one, it would be quite misleading to paraphrase his belief that Evans is a woman by saying that he believes Eliot to be a woman.

The constraints on legitimate paraphrase, and, in particular, the role that background knowledge plays in the matter, are, I suspect, complex. But this is not the place to go into them. It is clear that Kripke's problem arises because of a violation of these constraints. The contradictions that we are concerned with do not depend in any way on paraphrase; and SI is central to them.

2.7 Frege and SI

That SI fails in epistemic contexts, even when names are involved, is, of course, a well-known view. It was Frege's.[8] According to Frege's view, the Hooded Man inference fails because in the sentence 'You know that Cain was born in Megara,' 'Cain was born in Megara' refers not to its standard reference (which is, for Frege, a truth value), but to its sense, the thought (proposition) that Cain was born in Megara. And 'Cain' refers, not to its standard referent, the person, but to its standard sense, something like a conception of that person (an individual concept). Similarly, in the epistemic context, 'Nescio' refers not to the person but to a conception of a person. And even if Nescio and Cain are the same person, the two conceptions of the person are not. Hence, we cannot substitute the one for the other. (In a sense then, the failure of SI is merely syntactic, since we are not dealing with co-referring expressions.)

[8] See Frege (1952).

Unfortunately, Frege's account faces difficult problems.[9] Consider the inference:

> You know that Cain was born in Megara.
> Cain has red hair.
> ———————————————
> There is someone with red hair whom you know to have been born in Megara.

This would certainly seem to be valid, but it is not for Frege. Even if the premises are true, the conclusion:

(1) $\mathfrak{S}x\,(x$ has red hair $\wedge$ you know x was born in Megara).

is false. To make the sentence true, the first x has to be a person; the second x has to be an individual concept. And no person is an individual concept.

There are ways one might try to get around this problem. For example, one may bite the bullet and agree that (1) is really false. The truth that the conclusion is meant to express is:

> $\mathfrak{S}x\mathfrak{S}y\,(x$ has red hair $\wedge$ y is an individual concept of x $\wedge$ you know that y was born in Megara).

But if this is the conclusion of the argument, it would still seem to be invalid. For the conclusion now *entails* statements about individual concepts; but the premises certainly don't appear to do this.[10]

The difficulties with Frege's theory do not end here.[11] But the point of this book is not to discuss other views in detail: it is simply to paint one particular view. So, is there a semantics that preserves inferences such as Particular Generalization? As we will see, there is.

[9] Similar problems beset a paratactic account of *knowing that* (of the kind developed by Boër and Lycan (1986)) and any other account which makes it impossible for a variable to bind inside and outside an epistemic context simultaneously in the natural way.

[10] A similar argument is sometimes employed in connection with plural predication and quantification. There are some sentences that contain plural predicates and quantifiers, and that cannot be cashed out in terms of standard first-order quantifiers. A notorious example is: (2) There are some critics who admire only each other. Some writers have suggested that this sentence is actually a covert second-order sentence, quantifying over a set of critics. Such a suggestion would certainly appear to be incorrect. Assuming quantifiers to be existentially loaded, (2) appears to entail the existence of critics, but not of sets—which the second-order sentence does. See e.g. Yi (1999), esp. 165–6.

[11] Thus, for example, the whole idea that proper names have a semantically significant sense has been attacked by Kripke (1972).

2.8 SI and Open Worlds

One solution to the problem posed by SI[12] is to deploy open worlds in a slightly different way. In the semantics of these as specified, the sentence $n = c$ is treated as a predicate, '$x_1 = x_2$', to be satisfied by the denotations of the terms it contains, 'n' and 'c'. Suppose, however, that we treated it simply as a whole sentence. Thus, at each open world, δ will assign '$n = c$', and every other sentence, an extension and co-extension ($\{\langle\rangle\}$ or ϕ). The failure of the Hooded Man argument is then almost trivial. Here, for example, is a countermodel.

$\mathcal{C} = \{@\}$

$\mathcal{O} = \{w\}$

$D = \{0\}$

$@R^0_\Psi w$ (and only w)

$\delta(n) = \delta(c) = \delta(a) = 0$

$\delta^+(Mc, w) = \{\langle\rangle\}$

$\delta^+(Mn, w) = \phi$

It is simple to check that $@ \Vdash^+ n = c$, $@ \Vdash^+ a\Psi Mc$, but $@ \nVdash^+ a\Psi Mn$.

The trouble with this approach is evident enough, though. It does not extend to formulas that contain free variables in the scope of intentional operators. Thus, we lose the power to quantify into intentional contexts, and therefore the validity of intuitively perfectly correct inferences such as: $a\Psi Mc \models \mathfrak{S}x\, a\Psi Mx$ (e.g. you believe that Cain is your brother, so someone is believed by you to be your brother).

We could, of course, retreat from an objectual understanding of the quantifiers to a substitutional one. That would make the inference valid. But such a retreat is not congenial to noneism: a noneist accepts objects of thought as genuine, not just as linguistic simulacra. Treating quantification into intentional contexts as substitutional also forces us to treat quantification into ordinary contexts as substitutional. Thus, consider the inference 'Cain was born in Megara and you believe that Cain was born in Megara; hence someone is such that they were born in Megara and you believe them to have been so born':

$$\frac{Mc \wedge a\Psi Mc}{\mathfrak{S}x(Mx \wedge a\Psi Mx)}$$

[12] Sketched in Priest (2002).

This is intuitively valid; but the only way to make it so is to treat the quantifier as substitutional. Even if we were forced into substitutional quantification in intentional contexts, this should not disrupt objectional quantification in normal contexts.

Finally, and in any case, there would appear to be cases of true sentences where one quantifies into an intentional context, but where no instantiation is to be found. Consider, for example, that eerie feeling that we all of us have sometimes to the effect that something is wrong, but where one cannot put one's finger on what it is. Something is such that one knows that it is wrong, but there is no *c* such that one knows that *c* is wrong. Note that it would be wrong to say that the quantifier in question has narrow scope: one knows that something is wrong. Doubtless, this is true too. But in the case at point, one knows more than this. Something in the environment is wrong: a clock has been moved; the radio is missing. It is just that one cannot say what it is. Similarly, most people have experienced the feeling that there is something that they want to eat—an apple?, a pear?—but cannot quite put their finger on it—until the *aha!* moment: it was tomato.

2.9 Worlds and Identities

How else, then, might one proceed? Come back to the Hooded Man, and consider the situation concerning Cain and Nescio as you know it. These two might have the same identity; they might not. That is, there are worlds compatible with all that you know in which they do have the same identity, and worlds in which they do not. In particular, then, objects may have different identities at different worlds. For any object, then, there is a function that maps it to its identity at each world. Indeed, for technical simplicity, we can just identify the object with this function. Thus, we can take an object to be a map from worlds to identities.

Formally, the semantics look like this.[13] An interpretation is a structure $\langle \mathcal{P}, \mathcal{I}, \mathcal{O}, @, D, Q, \delta \rangle$. $\mathcal{P}, \mathcal{I}, \mathcal{O}$ and $@$ are as before. Q is a set of things that

[13] A similar semantics for 'contingent identity systems' can be found in Parks (1974). Other systems for contingent identity can be found in Hughes and Cresswell (1968), 198–9, Bressan (1972), and Gupta (1980). For a discussion of the first two of these, see Parks and Smith (1974), and Parks (1976), respectively. None of these is concerned with intentional contexts.

we may think of as identities. D is a collection of functions from worlds to Q; so that if $d \in D$, $d(w)$ is the identity of d at w.[14] δ assigns every constant a world-invariant denotation in D, and every n-place function symbol an n-place function on D. δ also assigns each predicate an appropriate extension and co-extension at each world in $\mathcal{C}$; ditto every matrix of the form $A \rightarrow B$, $\Box A$, and $\Diamond A$ at an impossible world, and every matrix at an open world. But the extensions and co-extensions are now subsets of (n-tuples of) Q, *not* D. In particular, if w is a possible world, $\delta^+(=, w) = \{\langle q, q\rangle : q \in Q\}$, and $\delta^-(=, w)$ is the complement of $\delta^+(=, w)$. Given an assignment of denotations (members of D) to the variables, a denotation is assigned to all terms just as before.

The denotation of any term is a function from worlds to identities. The appropriate truth/falsity conditions for atomic formulas at a world $w \in \mathcal{C}$ are therefore of the form:

$$w \Vdash_s^{\pm} Pt_1 \ldots t_n \text{ iff } \langle \delta_s(t_1)(w), \ldots, \delta_s(t_1)(w)\rangle \in \delta^{\pm}(P, w)$$

The truth/falsity conditions for formulas treated as atomic at impossible and open worlds are formulated in the same way, writing $\delta_s(t_i)(w)$ where before we wrote $\delta_s(t_i)$. The truth/falsity conditions for connectives and quantifiers are exactly the same as before; the propositional/ quantificational logic of these semantics is still, therefore, the same. In particular, all the standard quantificational rules, such as Particular Generalization, are valid.

These semantics invalidate SI in the context of intentional operators, as we shall see. But they invalidate it in any other context that involves more than one world. Thus, as the semantics stand, they also invalidate these inferences:

1. $a = b, \Box a = a \vdash \Box a = b$
2. $a = b, Qa \rightarrow Qa \vdash Qa \rightarrow Qb$
3. $a = b, fa = fa \vdash fa = fb$

[14] It is tempting to think of identities as Fregean senses, but this would not be right. If anything, it is the members of D that are more like senses, since they determine behaviour across worlds. This is essentially how members of D are, in fact, interpreted by Bressan (1972), Gupta (1980), and, in a similar semantics, Hintikka (1969). It would also be a mistake to interpret members of D, in the present semantics, as senses, however. They are simply the objects themselves.

To see this, choose an interpretation where $\delta(a) = d_a \neq d_b = \delta(b)$, and $d_a(@) = d_b(@)$. Then the premises are satisfied. (The second, in each case, is a logical truth.) For 1, let $d_a(w) \neq d_b(w)$, for some $w \in \mathcal{P}$. For 2, let $d_a(w) \neq d_b(w)$, $d_a(w) \in \delta^+(Q, w)$ and $d_b(w) \notin \delta^+(Q, w)$, for some $w \in \mathcal{C}$. For 3, let $\delta(f)(d_a) = e_a$, $\delta(f)(d_b) = e_b$, where $e_a(@) \neq e_b(@)$.

It is natural—though perhaps contentious[15]—to suppose that SI fails only in intentional contexts. Failures of the above kind can be ruled out by imposing these conditions:

(†) if $w \in \mathcal{C}$, $d(w) = d(@)$

(††) if for all $1 \leq i \leq n$, $d_i(@) = e_i(@)$
then $\delta(f)(d_1, \ldots, d_n) = \delta(f)(e_1, \ldots, e_n)$

The first condition is to the effect that variation of identity occurs only in open worlds: at all other worlds, the identity of an object is maintained. The second condition is to the effect that the functions that are the denotations of function symbols do not discriminate between things that have the same identities at @ (and so, by (†), at all worlds in $\mathcal{C}$). In other words, these conditions enforce the thought that variation of identity plays a significant role only at open worlds; that is, those worlds that represent certain intentional situations. We henceforth make these two assumptions, which entail that the only violations of SI occur within intentional contexts. (See the proof in the appendix.)[16]

Even with these restrictions, the semantics still invalidate SI in intentional contexts. The solution to the Hooded Man problem, in particular, is now simple. The inference in question ($n = c, a\Psi Mc \vdash a\Psi Mn$ is invalid.

[15] Thus, before the work of Kripke, few people subscribed to the validty of 1; and we will see some reason to doubt the validity of 3 in Ch. 8.

[16] It might be thought that no variation of identity across worlds can obtain, for the following reason. Since @ is the actual world, one might expect to have, for any predicate, P, and closed terms $t_1, \ldots, t_n$: ($\star$) $Pt_1 \ldots t_n \Leftrightarrow @ \Vdash^+ Pt_1 \ldots t_n$. Now we can reason: Let a and b be any objects. Suppose that $a(@) = b(@)$. Then $@ \Vdash^+ a = b$ (by truth conditions), and so $a = b$, by ($\star$). Hence, for all w, $a(w) = b(w)$ (by SI). But ($\star$) is, in fact, false. What the actuality of @ actually delivers is: ($\star\star$) $Pt_1(@) \ldots t_n(@) \Leftrightarrow @ \Vdash^+ Pt_1 \ldots t_n$. In a similar way, suppose that one is a four-dimentionalist about objects over time. Then 'John is happy' holds at time t iff *the part of John at t* is happy. John himself is a space-time worm, and not the sort of thing that is happy or not. The deviation from homophony is required because each sentence is, effectively, an indexical one, but we are giving non-indexical truth conditions.

To see this, take the interpretation where:

$$\mathcal{C} = \{@\}$$
$$\mathcal{O} = \{w\}$$
$$Q = \{0, 1\}$$
$$D = \{d_1, d_2\}, \text{ where } d_1(@) = d_2(@) = d_1(w) = 0 \text{ and } d_2(w) = 1$$
$$@R_{\Psi}^{d_1} w \text{ (and only } w\text{)}$$
$$\delta(a) = \delta(c) = d_1, \delta(n) = d_2$$
$$\delta^{+}(Mx, w) = \{0\}$$
$$\delta^{-}(Mx, w) = \{1\}$$

(Mx being a matrix). For future reference, call this interpretation $\mathfrak{I}$. $d_1(@) = d_2(@)$, so $@ \Vdash^{+} c = n$. And since $d_1(w) = 0 \in \delta^{+}(Mx, w)$, $w \Vdash^{+} Mc$. Hence, $@ \Vdash^{+} a\Psi Mc$. But since $d_2(w) = 1 \notin \delta^{+}(Mx, w)$, $w \nVdash^{+} Mn$, so $@ \nVdash^{+} a\Psi Mn$. (Indeed, $@ \Vdash^{+} \neg a\Psi Mn$.)[17]

Let me finish this section by clearing up a couple of possible misunderstandings about these semantics. One might be inclined to think that in these semantics terms are not really rigid designators. After all, the bit of them that is doing the work at each world, so to speak, varies from world to world. But this would be a misperception. The object denoted by a term does not vary from world to world. Specifically, the denotation function for terms carries no world-parameter, as that for predicates does, and as it would do in a semantics for non-rigid designators. Moreover, the semantics verifies quantifier principles that hold only for rigid designators. Thus, consider the inference $a\Psi\, b = b \vdash \mathfrak{S}x\, a\Psi\, b = x$. Like all instances of Particular Generalization, this is valid. But it would not be if the denotation of b varied from world to world: it might not be the same x that is identical to b at each world accessed by $R_{\Psi}^{\delta(a)}$. In the same way, let definite descriptions be understood in a non-rigid way. Then it is necessarily the case that the person who solved Fermat's Last Theorem is the person who solved Fermat's Last Theorem. But it is not the case that there is some person such that necessarily they solved Fermat's Last Theorem.

17 Kripke's puzzle about belief is also solved. Let Ψ, M, p, m, and g be: 'believes that', 'is a man', 'Pierre', 'Marian Evans', and 'George Eliot', respectively. Then it is easy enough to construct models where $g = m$, $p\Psi Mg$ and $p\Psi\neg Mm$ all hold. Salmon (1995), 5, points out that there may well be another rigid designator, n, such that $n = m = g$ and Pierre has no beliefs about Mn at all. It is easy enough to construct models where, in addition, both $\neg p\Psi Mn$ and $\neg p\Psi\neg Mn$ also hold.

It might also be tempting to think of the values of a function, d, in D as the *parts* of the object d at each world. In the same way, if this were a temporal logic, it would be natural to think of d as an object comprising temporal parts, and as the members of Q as the temporal parts. And one can certainly conceptualize things in this way. I think that this is the wrong way to think about things, at least in the intentional case, however. For a start, identities vary only at open worlds, whereas temporal parts change at all worlds (= times). But even if we were to allow identities to change arbitrarily, this would still, I think, not be the best way to look at things. To do so would be to take the parts to be metaphysically primary, and an object to be the sum of its parts. I think that it is preferable to take the members of D to be metaphysically primary. The value of an object, d, at each world is its identity there. At each world an object has an identity, just as much as it has a length, a colour, and so on. (All the worlds are stages, and all the people merely players.)

One argument for this is as follows. If there were parts that were metaphysically primary, there would appear to be no reason why every function from worlds to parts should not constitute an individual. (There are no privileged linkages between world parts.) But if this were the case, as is shown in the appendix, the following would be a valid inference: $a\Psi\mathfrak{S}xMx \vdash \mathfrak{S}x\, a\Psi Mx$. But this is certainly not valid. I can know that there are spies without knowing of any person that they are a spy.[18]

2.10 The *De Re* Argument

We are still not finished with the Hooded Man argument. There is a distinction, dating back to Medieval Logic, that is standardly drawn between two different understandings of a statement of the form 'It is known

[18] Finally, one might suppose that these semantics are just Lewis's modal counterpart semantics ((1968)—page references to the reprint) in disguise. Thus, it might be thought that the functions in D are simply the counterpart relations between members of Q. Specifically, we might think that x is a counterpart of y iff: for some $d \in D$ and $w_1, w_2 \in \mathcal{W}$, $d(w_1) = x$ and $(w_2) = y$. This, however, is not the case. First, there are questions of interpretation. In counterpart theory it is the members of Q that are the genuine objects, and so constitute the domain of quantification; in the above semantics it is the members of D that are the genuine objects. Secondly, there are differences between the properties of the above relation and a counterpart relation. The above relation is clearly symmetric, but a counterpart relation need not be (p. 28–9). Third, these differences affect the resulting logic. For example, the universal closure of $A \supset \Box\Diamond A$ (valid in $S5$) fails in counterpart theory, since the counterpart relation is not symmetric (p. 36); it is valid in the above semantics.

that Cain was born in Megara.' On the first understanding, *de dicto*, this expresses a property of a proposition, or some other kind of truth-bearer, such as a sentence; in this case, the proposition (or sentence) *Cain was born in Megara*. The intentional sentences we have been concerned with so far are, in fact, all of this kind.

On the second understanding, *de re*, the sentence is taken to express a predication of the object of the knowledge—in this case, Cain. The *de re* interpretation might be expressed more perspicuously as: Cain is such that you know him to have been born in Megara. It is usually claimed that SI holds for *de re* interpretations. Indeed, substitutivity is often taken as a criterion for being *de re*. Thus, there would appear to be another version of the paradoxical argument in the wings, which is as follows:

Nescio is Cain.
Cain is such that you know that he was born in Megara.

Nescio is such that you know that he was born in Megara.

Is the conclusion of this argument unacceptable though? Perhaps not. Nescio, *that very person*, *is* such that you know him to have been born in Megara. You just don't realize this.

But things are not that simple. Suppose that SI works in *de re* contexts. Then it is indeed true that Nescio is a person, namely Cain, such that you know him to have been born in Megara. But it would appear equally to be the case that Cain is a person, namely Nescio, such that you do not know him to have been born in Megara, since we have:

Nescio is Cain.
Nescio is such that you do not know him to have been born in Megara.

Cain is such that you do not know him to have been born in Megara.

Let us call this the *counter-argument*. It would seem to be just as good. And if so, there is a person (Nescio, i.e. Cain) such that you both know and do not know him to have been born in Megara. We still appear to have a contradiction on our hands. What is to be said of this?

One possible solution to the problem is to insist that the second premise of the counter-argument, that Nescio is such that you do not know him to have been born in Megara, is just false. He *is* such that you know him to have been born in Megara; you just do not realize this fact. You may not know, *de dicto*, that Nescio was born in Megara. But what you know about Nescio *de re* is not open to introspection, simply because you may not recognize him under certain guises.

This is certainly a possible solution, but it has its problems. We have granted the *de dicto* claim that you do not know that Nescio was born in Megara. Moreover, 'Nescio' here is a rigid designator. It refers to that very object, independently of how it is picked out in a particular world. There may even be a causal (indeed perceptual) baptism of Nescio with this name. It would seem to follow that the epistemic state is also *de re*. That is, *de dicto* + rigid designation + perceptual contact entails *de re*.[19]

Perhaps there are replies to this objection. The notion of *de re* knowledge is, after all, slippery enough. But is there another possible solution? There is. To see what it is, consider the sentence:

Cain is such that you know that he was born in Megara.

What is its logical form? The way to represent the sentence that sticks most closely to its surface form is obtained by employing λ-abstraction, so that it may be represented as $\lambda x(a\Psi Mx)c$.[20] But we can avoid introducing this new machinery. The point of a *de re* claim is that it is a claim about the object itself, independently of how it is referred to. And since reference to objects themselves is carried by quantifiers, we can capture the content of the claim by:

$$\mathfrak{S}x(x = \text{Cain} \wedge a\Psi\ x \text{ was born in Megara})$$

Hence, the *de re* inference is of the form:

$$\frac{\begin{array}{l} n = c \\ \mathfrak{S}x(x = c \wedge a\Psi Mx) \end{array}}{\mathfrak{S}x(x = n \wedge a\Psi Mx)}$$

This argument, involving substitution, as it does, only in non-intentional contexts, is valid. Given that the premises are true, we therefore accept the conclusion: the Hooded Man, Nescio, *is* such that you know him to

[19] More: suppose that five minutes after Nescio enters the room, someone reliably tells you of Nescio that he was, in fact, born in Megara. Your *de re* knowledge of Nescio would seem to have changed. Yet this cannot be the case if you already knew of Nescio that he was born in Megara.

[20] For an account of λ-terms in the context of quantified modal logic, see Fitting and Mendelsohn (1998), chs. 9, 10.

have been born in Megara. (Though, as referred to by the name 'Nescio', you may not realize this.)

What of the counter-argument? The logical form of the argument is:

$$\frac{\begin{array}{l} n = c \\ \mathfrak{S}x(x = n \wedge \neg a\Psi Mx) \end{array}}{\mathfrak{S}x(x = c \wedge \neg a\Psi Mx)}$$

and it, too, is valid. Hence, if the premises are true, so is the conclusion: Cain is such that you do not know him to have been born in Megara. (Though, as referred to by the name 'Cain', you may not realize this.)

Thus Nescio (that is, Cain), is such that you both do and do not know that he was born in Megara. This may *sound* like a contradiction, but it is not. It is of the form:

$$\mathfrak{S}x(x = n \wedge a\Psi Mx) \wedge \mathfrak{S}x(x = n \wedge \neg a\Psi Mx)$$

Of course, the x in question is n, and $\neg a\Psi Mn$; but any attempt to obtain $a\Psi Mn$, and thus an explicit contradiction, from the first conjunct falls foul of the failure of SI in intentional contexts. Indeed, the interpretation $\mathfrak{I}$ of the previous section makes this sentence true, since it makes $c = n \wedge a\Psi Mc$ and $n = n \wedge \neg a\Psi Mn$ true at @. The result follows by generalization.[21] Hence, the *de re* problem is also solved by the semantics.

2.11 Conclusion

Identity appears to be such a simple notion. Yet the journey concerning its behaviour in intentional contexts has been a quite long one. And we are not finished with identity yet: there will be more to say about it in Ch. 8. Yet we now, at least, have a semantics for identity that solves the Hooded Man Paradox—and the similar paradoxes for other intentional operators.

[21] If *de re* constructions are represented by λ-terms, we would have $\lambda x(a\Psi Mx)c \wedge \lambda x(\neg a\Psi Mx)c$—and the same for n. But this does not convert into a contradiction. λ-conversion will fail in epistemic contexts for exactly the same reason that substitutivity does.

This chapter and the last have given us a semantics for a quantified language with modal operators, a relevant conditional, identity, and intentional operators. We have said nothing yet about intentional predicates. To this subject we turn in the next chapter. The material there will be much simpler technically (due, in part, to the fact that the discussion can piggy-back upon that of intentional operators). However, the topic will bring the philosophical issues concerning noneism directly to the surface.

2.12 Technical Appendix

In this appendix I will verify the technical claims made in this chapter. First, consider the semantics for identity of 2.2. Note that nothing in the treatment of identity in this section affects the proofs of Lemmas 1 and 2 in 1.9, which therefore continue to hold.

Corollary 4 *If t_1 and t_2 are free when substituted in $A(x)$, then: $t_1 = t_2, A(t_1) \models A(t_2)$.*

Proof Suppose that $@ \Vdash_s^+ t_1 = t_2$ and $@ \Vdash_s^+ A(t_1)$. Then $\delta_s(t_1) = \delta_s(t_2)$. Let this be d. By Lemma 2, $@ \Vdash_{s(x/d)}^+ A(x)$, and so $@ \Vdash_s^+ A(t_2)$, by Lemma 2 again. ∎

We now show that with the semantics of identity of 2.9, Lemmas 1 and 2 continue to hold. Specifically:

Lemma 5 *Fix any interpretation. Let t and A be any term and formula. Then if s_1 and s_2 are any evaluations of the variables that agree on the variables free in t and A:*

1. $\delta_{s_1}(t) = \delta_{s_2}(t)$
2. for all $w \in \mathcal{W}$, $w \Vdash_{s_1}^{\pm} A \Leftrightarrow w \Vdash_{s_2}^{\pm} A$

Proof The proof of 1 is exactly the same as that of 1 in Lemma 1. The proof of 2 is similar to that of 2 in Lemma 1. We simply replace everything of the form $\delta_{s_i}(t_j)$ in the atomic cases with $\delta_{s_i}(t_j)(w)$. ∎

Lemma 6 *Fix any interpretation. Let $t'(x)$ and $A(x)$ be any term and formula. Let t be any term that can be freely substituted for x in these. Let s be any* *evaluation of the free variables, then if* $d = \delta_s(t)$:

1. $\delta_{s(x/d)}(t'(x)) = \delta_s(t'(t))$
2. for all $w \in \mathcal{W}$, $w \Vdash^{\pm}_{s(x/d)} A(x) \Leftrightarrow w \Vdash^{\pm}_{s} A(t)$

Proof The proof of 1 is the same as that of 1 in Lemma 2. The proof of 2 is similar to that of 2 in Lemma 2. We simply replace everything of the forms $\delta_s(t_j(t))$ and $\delta_{s(x/d)}(t_j(x))$ in the atomic cases with $\delta_s(t_j(t))(w)$ and $\delta_{s(x/d)}(t_j(x))(w)$, respectively. ■

The correct behaviour of the quantifiers follows:

Corollary 7 *If t is free when substituted for x in* $A(x)$ *then*:

1. $\mathfrak{A}xA \models A(t)$
2. $A(t) \models \mathfrak{S}xA$

Proof As for Corollary 3 of 1.9. ■

As we saw in 2.9, we do not have substitutivity of identicals in general, but we do have it provided that we are not substituting into the scope of an intentional operator, Ψ. This follows from Lemma 8.

Lemma 8 *Fix any interpretation, with* $d, e \in D$. *Let* $d(@) = e(@)$. *Let* $t(x)$ *and* $A(x)$ *be any term and formula such that x is not in the scope of a* Ψ *in* $A(x)$. *Then for any evaluation, s*:

1. for all $w \in \mathcal{C}$, $\delta_{s(x/d)}(t(x))(w) = \delta_{s(x/e)}(t(x))(w)$
2. for all $w \in \mathcal{C}$, $w \Vdash^{\pm}_{s(x/d)} A(x) \Leftrightarrow w \Vdash^{\pm}_{s(x/e)} A(x)$

Proof The proof makes use of the conditions (†) and (††) of 2.9. It is worth noting that this is the only place in this appendix where these conditions are appealed to. The proof of 1 is by recursion on the way that terms are constructed. If $t(x)$ is a constant, c, or variable, y, other than x, the substitution is vacuous. So we have:

$$\begin{aligned}\delta_{s(x/d)}(c) &= \delta\ (c)\\ &= \delta_{s(x/e)}(c)\\ \delta_{s(x/d)}(y) &= s(y)\\ &= \delta_{s(x/e)}(y)\end{aligned}$$

The result follows in each case. For x itself:

$$\begin{aligned}\delta_{s(x/d)}(x)(w) &= d(w)\\ &= d(@) && \text{by } (\dagger)\\ &= e(@)\\ &= e(w) && \text{by } (\dagger)\\ &= \delta_{s(x/e)}(x)(w)\end{aligned}$$

For function symbols, $\delta_{s(x/d)}(ft_1(x)\ldots t_n(x))$:

$$\begin{aligned}&= \delta(f)(\delta_{s(x/d)}(t_1(x)),\ldots,\delta_{s(x/d)}(t_n(x)))\\ &= \delta(f)(\delta_{s(x/e)}(t_1(x)),\ldots,\delta_{s(x/e)}(t_n(x))) && \text{by IH and } (\dagger\dagger)\\ &= \delta_{s(x/e)}(ft_1(x)\ \ldots\ t_n(x))\end{aligned}$$

The result follows.

For 2, $A(x)$ is made up from atomic formulas and formulas that do not contain x free by means of $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\Box$, $\Diamond$, $\mathfrak{S}$, and $\mathfrak{A}$. The proof is by recursion. If A does not contain x free, then the result follows by Lemma 5. For atomic sentences, $w \Vdash^{\pm}_{s(x/d)} Pt_1(x)\ldots t_n(x)$:

$$\begin{aligned}&\Leftrightarrow \langle \delta_{s(x/d)}(t_1(x))(w),\ldots,\delta_{s(x/d)}(t_n(x))(w)\rangle \in \delta^{\pm}(P,w)\\ &\Leftrightarrow \langle \delta_{s(x/e)}(t_1(x))(w),\ldots,\delta_{s(x/e)}(t_n(x))(w)\rangle \in \delta^{\pm}(P,w) && \text{by 1}\\ &\Leftrightarrow w \Vdash^{\pm}_{s(x/e)} Pt_1(x)\ldots t_n(x)\end{aligned}$$

For $\neg$:

$$\begin{aligned}w \Vdash^{\pm}_{s(x/d)} \neg B(x) &\Leftrightarrow w \Vdash^{\mp}_{s(x/d)} B(x)\\ &\Leftrightarrow w \Vdash^{\mp}_{s(x/e)} B(x) && \text{by IH}\\ &\Leftrightarrow w \Vdash^{\pm}_{s(x/e)} \neg B(x)\end{aligned}$$

The cases for the other extensional connectives are similar.

For $\rightarrow$: if $w \in \mathcal{P}$, then $w \Vdash^{+}_{s(x/d)} A(x) \rightarrow B(x)$:

$$\begin{aligned}&\Leftrightarrow \text{for all } w' \in \mathcal{C}, \text{ if } w' \Vdash^{+}_{s(x/d)} A(x) \text{ then } w' \Vdash^{+}_{s(x/d)} B(x)\\ &\Leftrightarrow \text{for all } w' \in \mathcal{C}, \text{ if } w' \Vdash^{+}_{s(x/e)} A(x) \text{ then } w' \Vdash^{+}_{s(x/e)} B(x) && \text{by IH}\\ &\Leftrightarrow w \Vdash^{+}_{s} A(x) \rightarrow B(x)\end{aligned}$$

The case for falsity is similar. And if $w \in \mathcal{I}$, $A \to B$ is of the form $C(t_1, \ldots, t_n)$, where C is a matrix. Then $w \Vdash^{\pm}_{s(x/d)} C(t_1(x), \ldots, t_n(x))$:

$$\Leftrightarrow \langle \delta_{s(x/d)}(t_1(x))(w), \ldots, \delta_{s(x/d)}(t_n(x))(w)\rangle \in \delta^{\pm}(C, w)$$
$$\Leftrightarrow \langle \delta_{s(x/e)}(t_1(x))(w), \ldots, \delta_{s(x/e)}(t_n(x))(w)\rangle \in \delta^{\pm}(C, w) \text{ by } 1$$
$$\Leftrightarrow w \Vdash^{\pm}_{s(x/e)} C(t_1(x), \ldots, t_n(x))$$

For $\Box$: if $w \in \mathcal{P}$:

$$\begin{aligned} w \Vdash^{\pm}_{s(x/d)} \Box B(x) &\Leftrightarrow \text{for all/some } w' \in \mathcal{P}, w' \Vdash^{\pm}_{s(x/d)} B(x) \\ &\Leftrightarrow \text{for all/some } w' \in \mathcal{P}, w' \Vdash^{\pm}_{s(x/e)} B(x) \quad \text{by IH} \\ &\Leftrightarrow w \Vdash^{\pm}_{s(x/e)} \Box B(x) \end{aligned}$$

And if $w \in \mathcal{I}$, the argument is the same as that for $\to$. The argument for $\Diamond$ is similar.

Finally, for $\mathfrak{S}$, consider $\mathfrak{S}yB(x)$. If x is the same as y, then x is not free in $\mathfrak{S}yB(x)$, so the result holds by Lemma 5. So suppose that they are distinct:

$$\begin{aligned} w \Vdash^{\pm}_{s(x/d)} \mathfrak{S}yB(x) &\Leftrightarrow \text{for some/all } b \in D, w \Vdash^{\pm}_{s(x/d,y/b)} B(x) \\ &\Leftrightarrow \text{for some/all } b \in D, w \Vdash^{\pm}_{s(x/e,y/b)} B(x) \quad \text{by IH} \\ &\Leftrightarrow w \Vdash^{\pm}_{s(x/e)} \mathfrak{S}yB(x) \end{aligned}$$

The case for $\mathfrak{A}$ is similar. ■

Corollary 9 *If t_1 and t_2, when substituted for x in $A(x)$, are both free and not in the scope of an intentional operator, Ψ, then: $t_1 = t_2, A(t_1) \models A(t_2)$.*

Proof Suppose that $@ \Vdash^{+}_{s} t_1 = t_2$. Let $d = \delta_s(t_1)$ and $e = \delta_s(t_2)$. Then $d(@) = \delta_s(t_1)(@) = \delta_s(t_2)(@)) = e(@)$. So:

$$\begin{aligned} @ \Vdash^{+}_{s} A(t_1) &\Leftrightarrow @ \Vdash^{+}_{s(x/d)} A(x) \quad \text{by Lemma 6} \\ &\Leftrightarrow @ \Vdash^{+}_{s(x/e)} A(x) \quad \text{by Lemma 8} \\ &\Leftrightarrow @ \Vdash^{+}_{s} A(t_2) \quad \text{by Lemma 6} \end{aligned}$$

■

Finally, as promised:

Lemma 10 *If, in an interpretation, for any way of selecting a member of Q at a world, w, there is a $d \in D$ such that $d(w)$ is that member of Q, if $@ \Vdash^{+}_{s} a\Psi\mathfrak{S}xMx$ then $@ \Vdash^{+}_{s} \mathfrak{S}x\, a\Psi Mx$.*

Proof Suppose that $@ \Vdash_s^+ a\Psi\mathfrak{S}xMx$. Then for all w such that $@R_\Psi^{\delta_s(a)}w$, $w \Vdash_s^+ \mathfrak{S}xMx$; so there is a $d \in D$ such that $w \Vdash_{s(x/d)}^+ Mx$; that is, $d(w) \in \delta(M, w)$ (or $\delta(Mx, w)$ if $w \in \mathcal{O}$, where Mx is a matrix—take this qualification as read in what follows); so there is a $q \in Q$ such that $q \in \delta(M, w)$. Now for every such world, w, choose one such q, q_w (by the Axiom of Choice), and let $d^* \in D$ be such that $d^*(w) = q_w$; for other worlds, w, $d^*(w)$ can be anything one likes. Then for all w such that $@R_\Psi^{\delta_s(a)}w$, $d^*(w) \in \delta(M, w)$; i.e. $w \Vdash_{s(x/d^*)}^+ Mx$. So, $@ \Vdash_{s(x/d^*)}^+ a\Psi Mx$, and $@ \Vdash_s^+ \mathfrak{S}xa\Psi Mx$. ∎

3

Objects of Thought

3.1 Introduction: Intentional Predicates

In this chapter we move from intentional operators to intentional predicates, that is, intentional verbs whose complements are not whole sentences but noun-phrases, as in: 'Homer worshipped Zeus,' 'I am reading about Sherlock Holmes,' 'Alchemists sought the Philosophers' Stone,' and so on. I will give a semantics for such predicates and discuss some of its features.

Intentional predicates are involved in problems and apparent paradoxes just as much as intentional operators. We will approach the subject via three such. Discussions of these can certainly be found in modern philosophy, but some of the most sophisticated discussion of them occurred in medieval logic. I will therefore take the paradoxes from there. In particular, we will look at three of the sophismata of the great fourteenth-century logician, Jean Buridan.

For the purposes of the book, it is not necessary to know what the medieval logicians themselves thought about these problems. But I think that it is illuminating to know this—if only to remove the blinkers put on by a contemporary education in philosophical logic. The matter is discussed in the appendix to this chapter.[1]

3.2 Non-Existence

The three sophismata that we will take as our guide point to three apparent features of intentional predicates: these concern non-existence,

[1] This chapter of the book is written jointly with Stephen Read. The body of the chapter owes more to me than to him; the appendix owes more to him than to me. But it would be wrong to suppose that either of us is fully responsible for either part.

the failure of substitutivity, and indeterminacy. The phenomena are often run together under the blanket rubric of intensionality; but as we will see, they are quite distinct.

The first of these topics, and perhaps the most fundamental, is that of non-existence. On the surface, it is natural to parse a sentence such as 'Homer worshipped Zeus' as a binary relation between two objects, the cognitive agent in question, Homer, and the object of their intentional state, Zeus. But as the example shows, that object may not exist. How can there be a relationship between two objects, one of which does not exist? Here is the matter, as put by Buridan.[2] Note that each of Buridan's sophismata starts with a sentence, then gives an argument for the truth of this, and then an argument for its falsity. A commentary subsequently goes on to resolve the apparent contradiction.

Sophism: A non-being is understood.

Posit that the proposition is affirmative with an infinite subject. Then the sophism is proved: for such infinite terms are analysed so that to say 'A non-man runs' is equivalent to saying 'What is not a man runs'. And thus to say 'A non-being is understood' is equivalent to saying 'What is not a being is understood'. But the second is true, for Antichrist, who is not a being, is understood.

The opposite is argued; for the term 'non-being' supposits for nothing, but a proposition is false if its subject supposits for nothing and if it is affirmative; therefore, etc.

Buridan sets up the problem employing the categories of medieval logic. The central ones are explained in the appendix to this chapter, but one does not need a grasp of these to see Buridan's point. We understand the Antichrist. But how can that be, since the Antichrist does not exist?[3]

The solution to Buridan's problem to be endorsed here is simply to accept that an agent can have a relationship with a non-existent object: noneism. To suppose otherwise is simply a prejudice in favour of the actual, as Meinong put it. By analogy with 'racism' and 'sexism', etc., we might call this 'actualism'—though that word has other well-known uses in philosophy.

The noneist strategy is a very natural one. Thus, for example, when one fears something, one has a direct phenomenological experience of

[2] Buridan (2001), 923. See also Scott (1977), 97.

[3] Buridan, being a good medieval Christian, thinks that the Antichrist *will* exist. But future existence is not existence. For further details, see the appendix to this chapter.

a relation to the object of the fear. And the phenomenology is quite independent of whether or not the object *actually* exists. What more appropriate, then, to suppose that objects may exist or not, and that their existential status is irrelevant to whether or not they can be the target of an intentional state? The noneist generosity extends, note, just as much to impossible objects as possible objects. For one can think of the greatest prime number just as much as one can think of the smallest. And one can seek both a proof of Goldbach's Conjecture and a proof of its negation—though one of these cannot exist. An intentional predicate, then, is a relation that may be towards non-being.

This analysis of the objects of intentional predicates was advocated, perhaps most famously, by Meinong (though the view is certainly not original to him, as the appendix to this chapter shows). But since Meinong, many philosophers—starting with Russell—have felt that there is something philosophically rebarbative about the very notion of a non-existent object, that it is beset with insurmountable objections. Most of the objections to meinongianism have been demolished by various authors.[4] We will have occasion to look at rebuttals of some of the more major objections in due course.

Moreover, other attempts to give predicates such as 'fears' a parsing that does not invoke non-existent objects face well-known objections. The most common suggestion is to reparse the relation 'x fears y' (xFy)[5] as one between the agent and some surrogate object, especially some mental representation. Thus, 'Benny fears the man next door' might be understood as 'Benny has a man-next-door-representation in his mental "fear box" '. Call this relation F'. This suggestion won't do as it stands. For example, it makes a nonsense of 'There is someone whom Benny fears, but who is, in fact, a very nice man.' This is of the form: $\mathfrak{S}x(bF'x \wedge Mx)$. To make sense of the first conjunct, the quantifier has to range over representations, but then the second conjunct is nonsense: the representation is not a man at all—nice or otherwise.

One can save this view by invoking the relation 'x is a representation of y' (xRy). The sentence then becomes $\mathfrak{S}x\mathfrak{S}y(bF'x \wedge xRy \wedge My)$. But this is just the start of the problems. For example, how is one to understand 'Benny and Penny fear something (the same thing)'? $\mathfrak{S}x(bF'x \wedge pF'x)$ won't do: there is no guarantee that Benny and Penny have exactly

[4] Notably, Routley (1980), esp. chs. 3 and 4.

[5] As with identity, I will normally write binary intentional predicates between their arguments, rather than before them. This makes reading easier.

the same mental representation of the object in question. We can try $\mathfrak{S}x\mathfrak{S}y\mathfrak{S}z(bF'x \wedge pF'y \wedge xRz \wedge yRz)$. This may work if they both fear the same existent object (z). But the object in question may not exist; and if this is so, non-existent objects are still being invoked.

The natural thought at this point is to define an equivalence relation between representations, $\backsim$, such that $x \backsim y$ iff x and y are representations that appear to be of the same thing, z—and if z exists, actually are. We can then parse the sentence as: $\mathfrak{S}x\mathfrak{S}y(bF'x \wedge pF'y \wedge x \backsim y)$. The trouble, now, is with the relationship $\backsim$. Without noneism, it would seem hard to understand it. Different representations of the same object can be arbitrarily different. (Consider, for example, a painting of a place in the style of Constable, and another in the style of Picasso. For good measure, consider a musical representation of the same place in the style of Sibelius, and a digitalization of this.) Thus, the relation cannot be defined in terms of the *intrinsic* qualities of the representations. It can be done only in terms of their extrinsic qualities; and the most obvious of these, namely their representing a certain object, is ruled out unless one is a noneist (in which case it is unnecessary—for present purposes anyway).

Doubtless, there is much more to be said about all this; however, I will not go into it here. What follows will explore and develop the simple and natural noneist strategy.

3.3 A Formal Semantics

For a start, let me give a noneist semantics for intentional predicates. In fact, we can simply use the semantics already given for intentional operators, and take intentional predicates to be ordinary predicates. Things could not be easier! The domain in question must now, of course, contain objects that exist (at a given world) and objects that do not exist (at a given world)—as flagged in 1.4. As also flagged there, the quantifiers range over all the objects in the domain, existent and non-existent; and existence (at a world) is expressed by the monadic existence predicate.

Just as we may wish to add further constraints on the semantics for particular operators (see 1.3), we may wish to add further constraints for particular predicates. Thus, some predicates are existence-entailing. If a kicks b, or holds b, or runs past b, then both a and b must exist.[6]

[6] Consider the predicates 'x is transparent' and 'x is opaque'. These are both existence-entailing. Hence, if x does not exist, 'x is transparent' and 'x is opaque' are both false. What, then, makes them contraries? The fact that they cannot both be false for existent objects. Thus,

Some predicates are existence-entailing in some arguments, but not others. Intentional predicates are a prime examples of this. Thus, if a fears b, or thinks of b, or worships b, then a must exist but b may or may not.[7] It is clear, however, that some non-intentional predicates are not existence-entailing. Thus, logical predicates, such as identity, are not: even if a does not exist, it is still true that a is self-identical, $a = a$. (We will have some more examples of non-intentional non-(existence-entailing) predicates in Ch. 7.)

Of course, for certain predicates, it may be a matter of debate whether they, or some of their places, are existence-entailing. Thus, if a experiences b, and b is a visual after-image, does it follow that b exists? My aim here is not to enter into debates concerning this, or any other particular predicates. However, if P is a predicate that is existence-entailing in its ith place, the semantics needs to augmented by the constraint:

$$\text{if } \langle q_1, \ldots, q_i, \ldots, q_n \rangle \in \delta^+(P, @) \text{ then } \mathrm{q_i} \in \delta^+(\mathrm{E}, @)$$

This constraint will apply, in particular, to the first place of every intentional predicate (but not to the other(s)).

Note that the constraint applies only at the actual world. What happens at other worlds is another matter. As a matter of fact, it seems to me that existence-entailing is world-invariant, at least at possible worlds. Thus, for example, if a hits b at such a world, w, then a and b exist at w. But we do not need to take a stand on this matter here.

As is clear, given the semantics, objects may be either existent or non-existent (at a world). There are no constraints on existent objects: they can satisfy any predicate. Non-existent objects can satisfy (at least) the places of intentional predicates other than the first and logical predicates. Thus, for example, it is true that Homer worshipped Zeus; and it is easy to construct an interpretation where $@ \Vdash^+ hWz \wedge \neg Ez$. (In the interests of perspicuity, 'z' is a constant here.) They can also satisfy sentences of the form $\Diamond A(x)$ and $a\Psi A(x)$. (Thus, for example, it is easy to construct a model where h does not exist, but could do: $@ \Vdash^+ \neg Eh \wedge \Diamond Eh$.) But, trivially,

we do not have $\mathfrak{A}x$(x is transparent iff x is not opaque). But we do have $\forall x$(x is transparent iff x is not opaque).

[7] It is not uncommon to find philosophers (not to mention any names!) arguing that intentional relations are not really relations, since relations require the existence of their relata, demonstrating this last claim by taking an existence-entailing relation, such as 'x hit y', and pointing out that if x hit y then x and y exist. The invalidity of inferring a property of all relations from the fact that one relation has it is staggering.

the semantics rules out non-existent objects having existence-entailing properties.

3.4 Substitutivity of Identicals

Let us turn from the first feature of intentional predicates to the second. This is the apparent failure of substitutivity. Thus, let us suppose that I love Jezebel. Jezebel, unbeknownst to me, is the most evil woman in the world. It would not seem to be true that I love the most evil woman in the world. Here is Buridan again:[8]

Sophism You know the one approaching.

I posit the case that you see your father approaching from afar, so that you cannot tell whether he is your father or someone else. Then the sophism is proved as follows: you know your father well; and your father is the one approaching; therefore you know the one approaching.

Again, you know the one who is known to you; but the one approaching is known by you; therefore you know the one approaching.

I prove the minor: for your father is known by you, and your father is approaching; therefore, etc.

The opposite is argued: you do not know the person concerned when he is such that, if asked who he is, you would truly say: 'I do not know'; but about the one approaching you will say this; therefore, etc.

As is evident, this is a version of the Hooded Man argument. But the intentional context involved here is clearly a predicate, not an operator. There are, of course, many similar examples. Oedipus desired Jocasta; but it does not seem to be the case that he desired his mother—though Jocasta was, unbeknownst to him, his mother. Or someone may fear Jack the Ripper, though not their next-door neighbour—even though the two persons are, in fact, the same.

Note that the problem is quite independent of the problem of existence, and can arise whether or not the object in question exists. Thus, Buridan's example applies to existent objects, but the following inference also seems to fail: John is thinking about Sherlock Holmes, so John is thinking about the killer of the Hound of the Baskervilles. John, never having read that particular story, may not know that Holmes killed the Hound.

[8] Buridan (2001), 892–3. See also Scott (1977), 72.

The examples appear to demonstrate a failure of SI within the scope of intentional predicates. But given the semantics of the previous section, SI holds provided that substitution is not into the scope of an intentional operator. In particular, therefore, it holds for substitution into the scope of intentional predicates. These, after all, behave just like any other predicate in this regard. Thus, if P is any intentional predicate we have: $b = c, aPb \vDash aPc$.

The solution to Buridan's problem is simple, however. It is not so clear that the example-situations really are as described. We may insist, for example, that Oedipus *did* desire his mother. He just did not realize that the object of his desire was his mother. Of course, Oedipus realized that his mother was his mother, and Jocasta was his mother; but it does not follow that he realized that Jocasta was his mother. Such an inference involves substitution into the scope of an intentional operator; and it was precisely the force of the last chapter that this fails. Similarly, if I love Jezebel, who is the most evil woman in the world, I do love the most evil woman in the world, though I do not realize that she is the most evil woman in the world, and would, presumably, cease to love her if I found this out.

There might appear to be harder cases to deal with when we consider intentional predicates with more than two arguments. Thus, for example, Lois Lane preferred Superman to Clark Kent. But these two are one. It would not appear to be the case that she preferred Clark Kent to Superman as, say, the girl who brings round the office doughnuts does—or even incoherently preferred Clark Kent to himself. However, these examples can be handled in the same way. Lois Lane really did have these preferences: she just did not realize that she did. Our appreciation of our own mental states is, after all, not incorrigible, especially when we are dealing with *de re* states, which these, in effect, are. So what is the difference between Lois Lane and the girl who brings round the office doughnuts? Simply that Lois Lane thinks that Superman is better than (preferable to) Clark Kent. For the office girl, it is the other way round. And since 'thinks that' is an intentional operator, we cannot transform one of these sentences into the other by substitution. Similarly, what is the difference between Lois Lane and the person who, incoherently, prefers Clark Kent to Clark Kent? Simply that Lois Lane does not know *that* Clark Kent is Clark Kent, whilst the person with incoherent preferences does: they prefer x to y even though they know that x is y.

Thus, the way is clear to accept substitution in intentional predicates, as the semantics deliver. Indeed, given the understanding of an intentional predicate as simply expressing a relation between two objects, it would seem that it must hold.

3.5 Indeterminacy

So let us turn to the third problem. This concerns the behaviour of quantifier phrases in the scope of an intentional verb. At first appearance, the matter here would seem to be straightforward. Thus, for example, consider: I fear every spider (or chimera). This is naturally represented as $\mathfrak{A}x(Sx \supset aFx)$. Or again: I fear no spiders (or chimeras). This is $\neg\mathfrak{S}x(Sx \wedge aFx)$.

But things are more complicated when we come to quantifier phrases of the form 'a/an *A*'. Suppose I am looking for a hotel. There may be no particular hotel that I am looking for, so it would be wrong to parse this as: $\mathfrak{S}x(Hx \wedge aLx)$. Here, again, is Buridan:[9]

Sophism 'I owe you a horse', and likewise, 'I owe you a penny'.

And I posit the case that in return for some good service that you performed for me, I promised you one good horse, and that I obligated myself before a competent judge to give you one good horse.

Then the sophism manifestly appears [to be true]: for it is commonly said that everything promised is something owed. And since this I owe, as long as I do not deliver what I obliged myself to deliver by means of a legitimate obligation before a judge, you can justly sue me in order that I deliver a horse to you, and this you could not do if I did not owe it to you . . .

But the opposite side is argued for in a way that is difficult to solve, granting the aforementioned cases, thus: nothing is owed by me to you; therefore I owe you neither a horse nor a penny.[10]

The consequent seems to be self-evident. For if you were to acknowledge before a judge that no thing is owed by me to you, the judge would rule that I was free from debt.

Thus the sentence:

(1) I promise you a penny.

[9] Buridan (2001), 907. See also Scott (1977), 83.

[10] Buridan's argument is rather cryptic at this point. What he means is that there is no penny (horse) such that I owe you that penny (horse). Hence, I do not owe you a penny (horse).

can be true even though there is no particular penny that I have promised you. I seem to have promised you some kind of indeterminate penny. Worse: for each penny, I have not promised you that penny. So how can I have promised you a penny?

Note that this problem, too, is quite independent of the problem of existence. The problem arises with this example, even though lots of pennies exist. Conversely, there is no indeterminacy in 'I seek Atlantis', even though Atlantis does not exist.

Towards a solution, start by noting that (1) (and any example of similar kind) is, in fact, ambiguous. It *can* mean that there is some particular penny that I have promised. For example, I may have promised you the first penny minted in England in 1900. On the other hand, there may be no particular penny: you lent me a penny, and I have simply promised to repay it. We can call these the *determinate* and *indeterminate* senses of (1) (and its kin), respectively. How can one tell the difference? In the determinate case, one can ask the question 'Which penny?' and expect to receive a sensible answer, such as 'the first penny minted in 1900'. In the indeterminate case, one cannot. If I lend you a penny, and you promise to give me a penny back at a later date, to ask 'which one?' would normally be a joke. There is no particular penny such that I have promised to give you *that* penny.[11]

Now, the determinate sense of (1) is easy enough. It is simply of the form:

(2) $\mathfrak{S}x(Qx \wedge aPx)$

(Here, Qx is 'x is a penny', and yPx is 'y promises you x'.) There is a wrinkle here, though. I may have promised you a particular penny, which is, in fact, fictional. Suppose, for example, I mistakenly take the works of Doyle to be history, rather than fiction; I may have promised you the last penny touched by Sherlock Holmes before he fell over the Reichenbach Falls. Being a penny is, arguably, existence-entailing. This object, being non-existent, is not, therefore a penny (at @). So the above representation is incorrect. I do, however, believe the object in question to be a penny.

[11] One might object: even in that case, there *is* some penny that I promised to give you: the penny that I promised to give you. But that can't be right. Suppose that I have two pennies. They can't *both* be the penny that I promised to give you, else I would have promised you twopence. Whichever isn't it, I couldn't then keep my promise by giving you that one. But that's silly.

So we can represent the situation as: $\mathfrak{S}x(a\Psi Qx \wedge aPx)$, where Ψ is an appropriate intentional operator. Or, to capture both situations under one general rubric, the sentence is: $\mathfrak{S}x(\,(Qx \vee a\Psi Qx) \wedge aPx)$.[12]

But what of the indeterminate sense of (1)? We should be clear, to start with, that noneism per se does not solve the problem of what this is. It might be thought to do so because, notoriously, non-existent objects are often claimed to be indeterminate in certain ways. Thus, for example, Meinong held that the Golden Mountain is neither rugged nor smooth, neither 15 carat nor 22 carat. We will come to this issue in a later chapter. Suppose, for the moment, that the thought is right. If I say 'I promise you a penny' in the indeterminate sense, maybe I promised you an indeterminate, non-existent penny? This thought does not survive long. I promised no such thing. If I had, it would make sense to ask 'which penny was promised?' and answer with 'a certain non-existent object'. But that is manifestly not what was promised. In the indeterminate case, the question makes no sense. We have, therefore, to look elsewhere.

Since the problem does not arise when what follows the verb is a proper name, but only when it is something of the form 'a so and so'; and since phrases of this kind often express particular quantification in English (e.g. in 'every man loves a woman'), an analysis in terms of quantifiers begs to be given. The trouble is that quantification doesn't seem to get us what we want. Writing (1) as (2) gives it the wrong sense, and because what follows the verb is not itself a sentence, there is nowhere else to insert the quantifier.[13]

[12] This dismantles a problem of Geach (1967). Geach worries about how to understand the sentence 'Hob thinks that a witch blighted Bob's mare, and Nob wonders whether she (the same witch) killed Cob's cow.' Since the pronoun 'she' picks up a reference to a particular witch, the question 'which witch?' makes sense. Hence, we should parse the sentence as: $\mathfrak{S}x$(x is (believed to be) a witch $\wedge$ Hob thinks that x blighted Bob's mare $\wedge$ Nob wonders whether x killed Cob's cow). Geach considers this suggestion (p. 148), and rejects it on the ground that it entails the existence of witches. Not if one is a noneist. A more difficult version of Geach's problem is given by Edelberg (1986), who sketches a noneist solution.

[13] It might be suggested that the sentence should not be understood as employing a quantifier, but as employing an indefinite description operator, ε, thus: I promise you εx(x is a penny). We will look at the semantics of such an operator in detail in the next chapter. As we shall see there, since there are lots of pennies, 'εx(x is a penny)' denotes one such, chosen non-deterministically. But even though the choice is non-deterministic, the description still denotes one particular penny; and in the indeterminate case, I did not promise you *that* particular penny—or I would not be able to repay you with some other penny, which I obviously can. So this suggestion will not work.

We may solve this problem as follows. When I say 'I promise you a penny', what I would normally, in fact, be doing is promising *to give* you a penny. If this is the case, we can analyse the content of what was said as:

(3) I promise that $\mathfrak{S}x$(x is a penny $\wedge$ I give you x).

Similar cases of indeterminacy can be handled in the same way. Thus, an utterance of 'I am looking for a hotel,' in the indeterminate sense, would normally mean 'I am trying to find a hotel,' i.e.: I am trying to bring it about that $\mathfrak{S}x$(x is a hotel $\wedge$ I find x).[14] Similarly, if I like a good curry then, normally, what I like is to eat a good curry. Thus we have: I like it to be the case that $\mathfrak{S}x$(x is a good curry $\wedge$ I eat x). Note that, in unusual circumstances, I might like the curry for some other purpose. For example, if I were a sexual pervert of a certain kind, I might like it to . . . with. In the appropriate context, then, the utterance could mean: I like it to be the case that $\mathfrak{S}x$(x is a good curry $\wedge$ I . . . x). The relevant predicate can be determined only in context. The situation with the other examples is the same.

This solves the problem of the indeterminate sense of (1) and its like by construing utterances of such sentences as elliptical for ones with a corresponding intentional operator; the indeterminacy is then handled by appropriately placing a particular quantifier. The indeterminate sense of (1) is, then, to be understood as of the form:

$$a\Psi\mathfrak{S}x(Qx \wedge aGx)$$

Notice how this explains the failure of the inference to the falsity of (1). For each penny, I did not promise (to give) you that penny: $\mathfrak{A}x(Qx \supset \neg a\Psi aGx)$. This clearly does not entail $\neg a\Psi\mathfrak{S}x(Qx \wedge aGx)$. (An interpretation to demonstrate this is left as a simple exercise.)

The crucial question is whether this strategy of turning a predicate into an operator is always available to us. When there is a case of indeterminacy, can the utterance always be taken as elliptical for one with an intentional operator? There certainly are uses of intentional verbs which resist being understood as expressing any kind of notion with a propositional complement. Thus, if I worship Zeus, this fact cannot

[14] It could be objected that if this were correct, it would follow that I am trying to bring it about that $\mathfrak{S}x$(x is a hotel)—which I am not. But as we know from Ch. 1, intentional states of this kind are not closed under entailment.

be cashed out as any particular intentional propositional attitude. Similarly, if I hallucinate a monster, there is no corresponding propositional state. Some writers, for example, Lakoff,[15] have mooted the possibility of there being covert such notions for which we currently have no name. Thus, for Lakoff, to admire x is to *wurf* to *glip* x. To endorse this view would, however, be an act of desperation. No content whatever can be given to wurfing or glipping. These are pseudo-notions. Hallucinating a monster is irredeemably hallucinating *something*, not F-ing *that* anything. It speaks in favour of the analysis, then, that with verbs that resist this kind of glossing, cases of indeterminacy do not seem to arise. If I say 'I worship a Greek god,' the question 'which one?' always seems to make sense. If I say 'I hallucinated a monster,' the question 'what was *it* like?' is always appropriate. Or consider Lakoff's example. 'Admire' is an intentional verb which, despite what he says, it seems impossible to paraphrase in a propositional fashion. 'I admire John' does not seem to be equivalent to anything of the form 'I . . . that . . . John . . .'. Now consider 'I admire a well-dressed woman.' This clearly has a determinate sense: there is some particular well-dressed woman I admire. In this case, there is even a universal sense: I admire any well-dressed women ($\mathfrak{A}x(x$ is a well-dressed woman $\supset$ I admire x)). But what there does not seem to be is any indeterminate sense.

It seems natural to conclude, therefore, that indeterminacy arises only when the statement made is, effectively, one with a that-clause. And if this is right, the solution sketched above is quite general.

3.6 Conclusion

We now have an account of the semantics of intentional predicates, and this has put noneism squarely on the agenda. I have appealed to the fact that one can quantify over, refer to, and individuate non-existent objects. The semantics given assumes, explicitly, that non-existent objects do not have existence-entailing properties; but what else is to be said about the properties of such objects? This raises the important issue of Characterization, the topic to which we will turn in the next chapter. Before we do this, however, and since the naturally curious reader will want to know what Buridan himself made of his sophisms, we will look in the following appendix at what he and some other medieval logicians

[15] Lakoff (1970), 221.

had to say on these matters. The appendix can be skipped without loss of continuity.

3.7 Appendix: Medieval Accounts of Intentionality

In this appendix we will look at the medieval views concerning the three sophismata that we have encountered—and especially the views of Buridan. For reasons of exegesis, we will take the topics in a slightly different order, starting with non-existence.

3.7.1 Non-Existence

Medieval logicians' accounts of intentionality piggy-back upon their general logical theory. Since this is unfamiliar to most contemporary logicians, let us start with a summary of its relevant features. Medieval logicians took simple sentences (i.e. those not containing connectives such as disjunction and the conditional) to be constituted by two terms related by the copula (hence the name for these logicians: 'terminists'), e.g. '*every person* is *one with a father*'. As the example illustrates, though, terms could be complex, and might be what we would now think of as quantifier expressions.

The terminists explained the semantics of such sentences by invoking various properties of the terms and of their parts.[16] One of these is *signification*. To one group of terminists, including William of Ockham, writing in the generation before Buridan, the signification of a term is simply its extension. Thus, 'penny' signifies pennies. To another group, this was too radical. For Buridan, for example, the concept F is abstracted from Fs by an act of mind and forms a natural likeness of them. English-speakers then adopt the convention of letting the sound 'penny' signify the concept *penny*. So by convention, the sound 'penny' ultimately signifies pennies via its immediate signification of the concept.

Next, we turn to the even more important notion of *supposition*. The supposition of a term is relative to the particular sentence in which it occurs. It is, again as a first cut, what the term refers to, as required by the truth conditions of the sentence. Thus, in 'Man is a species,' 'man' supposits for a universal; in 'Man has three letters,' it supposits

[16] See e.g. Read (2001). In 'Everyone has a father' the terms are 'everyone' and 'one with a father', but most medievals would also consider the properties of the sub-term 'a father' as well.

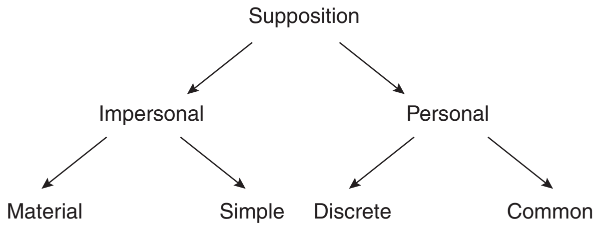

Figure 3.1

for a word. Cases of this kind were called *simple* and *material* supposition, respectively.[17] When a term supposits for what it signifies, it was said to have *personal* supposition. (At least, this is how both Ockham and Buridan used the word.) Personal supposition is itself of different kinds. When the term supposits for one particular object, in the way that a proper name or a definite description does, it was said to have *discrete* supposition. Otherwise it had *common* supposition, and supposited for a whole bunch of things. There were various distinctions drawn within common supposition, too. We will come back to these in due course. For present purposes, we can tabulate the distinctions within supposition as at Fig. 3.1.

The third property of terms to be noted is *ampliation*. Various verbs, or their features, may change the supposition-range of the terms in the sentence in which those verbs occur. For example, consider the sentence 'The Pope is walking.' In this, 'the Pope' has discrete supposition and supposits for a certain man, who now exists. But consider the sentence 'Plato walked.' Anyone who is dead no longer exists. Hence, there is nothing for the term 'Plato' to supposit for. To allow it to supposit, the tense of the verb 'walked' must allow the term 'Plato' to supposit not just for present objects, but for past objects too. This is ampliation. Similarly, 'The Antichrist is walking' is false, for the subject refers to nothing presently existing. But 'The Antichrist will walk' is true (according to the medievals), for the future tense of the verb ampliates the subject to present and future objects, and the Antichrist will exist (and walk) in the future.[18]

[17] In fact, Buridan conflated them, since simple supposition is really just material supposition for the mental word.

[18] Buridan (2001), 299. It is worth noting that Buridan himself was not prepared to call this behaviour 'ampliation' in the case of a term with descrete supposition. He did hold that the term behaves in the way described, however. See ibid. 918–19.

Constructions other than tense also have the power to ampliate. Thus, it is presumably true that the Third World War may start next year—however much we might hope that it will not. The modal auxiliary 'may' ampliates 'the Third World War' to supposit not only for present, past, and future things, but also for merely *possible* things. Other modal auxiliaries, like 'can' and 'must', do the same. Here is Buridan on the matter: 'A term put before the verb "can" ... is ampliated to stand for possible things, even if they do not and did not exist. Therefore the proposition "A golden mountain can be as large as Mont Ventoux" is true.'

It is worth noting that the medievals also recognized the operation inverse to ampliation, which restricts a range of supposition, rather than extending it. Thus, for example, they held that 'In my pocket' restricts the supposition of 'coin' in 'Every coin in my pocket is a penny' to supposit only for coins in my pocket.

After this prolegomenon, we come to intentional verbs. The medievals claimed that verbs of this kind also have the power to ampliate the supposition of terms following them. Thus, 'I understand the Antichrist' is true, since 'the Antichrist' supposits for a future entity due to the ampliation of 'understand'.[19] Moreover, such verbs may ampliate not just to past and future objects, but also to merely possible objects. For example, in 'I seek Atlantis,' 'seek' ampliates the supposition of 'Atlantis' so that it may refer to a possible but non-existent object.[20] William of Sherwood and other thirteenth-century figures speak quite unguardedly of terms ampliated to things that do not exist.[21] And here is Paul of Venice on the matter:[22] 'The absence of the signification of a term from reality does not prevent the term's suppositing for it.'

Medieval logicians, then, were quite happy to countenance non-existent objects.[23] The standard classes of object to which ampliation allowed access were the past, the future, and the possible. Did they also countenance impossible objects? Certainly not Buridan. Buridan's analysis of the sophism, 'Non-being is thought of' (which we noted in 3.2),

[19] Buridan (2001), 299.

[20] It should be noted that in these examples, the term whose supposition is ampliated is, strickly speaking, only a part of the predicate. Thus the proper predicate in 'I seek Atlantis' is 'seeker of Atlantis'. (I *am* a seeker of Atlantis.)

[21] See e.g. De Rijk (1982), 172.

[22] Paul of Venice (1978), 13.

[23] In this, in fact, they were just following Aristotle's lead: 'even non-existents can be signified by a name', *Posterior Analytics* 92^b29–30. (Translation from Ross 1928.)

makes it clear that he, at least, believed that everything that did not exist was at least possible. He writes:[24]

> The sophism ['A non-being is understood'] is false, for the term ['a non-being'] supposits for nothing. And this is clear in the following manner: the verb 'to understand' or 'to be understood' ampliates supposition to past and future, and even to all possible things. Therefore, if I say 'A being is understood', the term 'being' stands indifferently for every present or past or future or possible thing. But the rule is that an infinitizing negation added to a term removes its supposition for everything for which it supposited and makes it supposit for everything for which it did not supposit, if there are any such things. Therefore in the proposition 'A non-being is understood', the term 'non-being' does not stand for some present, nor for some past, nor for some future, nor for some possible being; therefore it supposits for nothing, and so the proposition is false.[25]

Other authors, in contrast, believed that verbs such as 'signify' and words such as 'intelligible' could ampliate terms to a fifth class of objects, beyond the standard four (what is, was, will be, or can be) namely, what can be imagined. Marsilius of Inghen, for example, writes:[26] 'Ampliation is the supposition of a term . . . for its significates which are or were, for those which are or will be, for those which are or can be, or for those which are or can be imagined.'

What, however, can be imagined? Marsilius certainly does not think that everything can be imagined. The void can be imagined because it can be created by the omnipotency of God. But a chimera may or may not be imaginable. The notion of a chimera may, in fact, be understood in two ways. A chimera may simply be something with an unnatural combination of parts (the head of a lion, the body of a goat, and the tail of a serpent); but it may also be something that has the essences of each of its parts, which is impossible (since the pertinent essences are incompatible).[27] Indeed, it is not uncommon for medieval writers to use the chimera as a standard example of an impossible object. At any rate,

[24] Buridan (2001), 923. See also Ebbesen (1986), 137: 'Buridan holds that the ampliative force of "opinabilis" [believable] does not extend to impossible entities.' Buridan is cited as saying: 'Every term which supposits, supposits for that which is or can be or has been or will be; but . . . it is impossible that a chimera can be, or can have been or can come to be . . . [Hence] "A chimera is thinkable" is false.'

[25] Thus, note, Buridan thinks that 'The Antichrist is understood' is true, but 'A non-being is understood' is false. The negation acts *after* the ampliation.

[26] Maierù (1972), 182; cf. Bos (1983), 103.

[27] See e.g. Bos (1983), 192; cf. Ashworth (1977), 62.

Marsilius thinks that a chimera, taken in the first way, is imaginable; but taken in the second way, since it is literally impossible, is not.[28]

Paul of Venice, however, is prepared to go further. For him, a chimera is indeed impossible.[29] None the less:[30] 'The verbs "think of", "imagine" and so on, both when they occur with an embedded clause and when they take a direct object [e.g. "I conjecture a chimera," "I imagine a vacuum"] always . . . cover not-being as well as being.' Indeed:[31] 'Although the significatum of the term "chimera" does not and could not exist in reality, still the term "chimera" supposits for something in the proposition . . . "A chimera is thought of", since it supposits for a chimera.'

Thus, there were at least some medieval logicians who were 'fully-fledged' noneists.[32]

3.7.2 *Ockham on Indeterminacy*

So much for the matter of non-existence.[33] Buridan's solutions to the other two problems appear to have been somewhat unusual by medieval standards—and also somewhat problematic. So, first, let us look at Ockham's solutions. His solution to the problem of indeterminacy, in particular, is much more orthodox in medieval terms. To explain this, more needs to be said about the modes of common personal supposition.

Common supposition was usually divided into *determinate* supposition and *confused* supposition. The second of these was split again into *confused and distributive* and *merely confused*. Thus, we may tabulate the distinctions drawn within personal supposition as at Fig. 3.2.

28 Cf. Ashworth (1977), 72.

29 Paul of Venice (1978), 254.

30 Paul of Venice (1981), 76.

31 Paul of Venice (1978), 13. See also Paul of Venice (1499), fo. 13^{vb}: 'A fourth way of responding is better: verbs like "is understood", "we believe", "signifies", "supposits" and so on ampliate their subject and predicate for present, past, future or imaginable things. So the proposition "A chimera is understood" should be analysed like this: "This is understood and this is or can be imagined to be a chimera".'

32 A generation earlier than Paul in Oxford, we also find Ralph Strode saying (Maierù 1972: 176): ' "supposits" is an ampliative term just like "signifies" . . . and so we must concede that "chimera" signifies something, even though what it signifies does not exist, and it supposits for something which nonetheless does not exist, just as I can think of or imagine what does not exist. Indeed, the term "chimera" supposits for something truly in such a proposition as "A chimera is believed in".' It is not clear from the context, though, which notion of chimera he is operating with.

33 It is worth noting that noneism did not die out between the medievals and Meinong. Another prominent exponent is Reid. See Routley (1980), 835–50, and Nichols (2002).

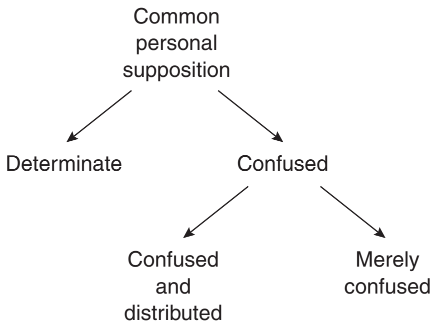

Figure 3.2

The marks showing that specific terms in a proposition have one of the three modes of common supposition were taken by Ockham, Buridan, and their followers to be the possibilities for descent from the said proposition to singular propositions (replacing the term and its determiner by a term with discrete supposition), and for ascent from those singular propositions to the original. Let us illustrate with examples.

Consider the sentence 'Some man is mortal.' One can infer from this that *this* man is mortal or *that* man is mortal or . . . , for an appropriate enumeration of men. Moreover, one can infer the sentence from each disjunct. This marks out 'man' as having determinate supposition here.

Next, consider the sentence 'All men are mortal.' One can infer from this that *this* man is mortal and *that* man is mortal and . . . , for an appropriate enumeration of men. One cannot, however, infer the sentence from any conjunct. This marks out the supposition of 'man' as confused and distributive here.

Finally, consider again 'All men are mortal,' but this time consider 'mortal'. One cannot infer 'All men are *this* mortal or all men are *that* mortal or . . .'. Nor can one infer the corresponding conjunction. But one can infer 'All men are *this* mortal or *that* mortal or . . .', for an appropriate enumeration of mortals. Moreover, one can infer 'All men are mortal' from 'All men are this mortal.' This marks out the supposition of 'mortal' as merely confused here.

We can now address the question of indeterminacy. Consider, again:

(1) I promise you a penny.

Ockham, Buridan, and their followers did not, in fact, accept that (1) is ambiguous. They insisted that it has only an indeterminate sense, the determinate sense being properly expressed by:

> (1a) A penny I promise you.

In this 'penny' has, happily enough, determinate supposition, since one can descend to:

> (4) *This* penny I promise you or *that* penny I promise you, and so on.

(and ascend from any disjunct).

For (1) itself, Ockham and Buridan part company. Let us follow Ockham for the present. According to him, the supposition of 'penny' in this is merely confused. One cannot infer the wide-scope disjunction, but one can infer:

> (5) I promise you *this* penny or *that* penny, and so on.

(and one can ascend from anything of the form 'I promise you this penny').[34]

Note that the disjunctions in (4) and (5) extend over all present and future pennies. 'Promise' is an intentional verb that ampliates the supposition of the term in such a way. For I can fulfil the promise by giving you a penny that does not yet exist. (In fact, medieval discussions of ampliation are more often carried out in the context of common supposition than discrete supposition. Thus, in 'All men will die,' the future tense ampliates the suppositional range of 'men', so that we can descend to '*This* man will die and *that* man will die and . . . , for all present and future men.[35])

Note, also, that since the term does not have determinate supposition, one cannot descend to 'I promise you this penny or I promise you that penny . . .' So, contraposing, one cannot ascend from 'I do not promise you this penny and I do not promise you that penny . . .' to 'It is not the case that I promise you a penny'. So the argument for the falsity of (1) is blocked.

[34] Ockham (1974), ch. 72, 207. Buridan (2001), 279, does not accept the descent as we shall see in a moment. Some logicians, such as William Burleigh, took 'a penny' to have simple supposition in (1). See Klima's introduction to Buridan (2001), lii.

[35] Ockham, in fact, analyses examples of this kind differently, by diagnosing an ambiguity, not by appealing to the notion of ampliation. See Priest and Read (1981).

The most notable difference between the Ockhamite account of the indeterminate case and the one offered in the chapter is that the former account does not require any propositional analysis of the indeterminate sense of (1) and its like. The indeterminate sense is obtained by attributing to 'promise' the power to cause terms following it to have merely confused supposition, just as it has the power to ampliate their supposition. This uniformity speaks in its favour. On the other hand, just because of the uniformity, one would have thought that it ought then to be possible to have an indeterminate sense in all cases. Thus, there ought to be an indeterminate sense of 'I worship a Greek god,' that is, 'I worship Zeus or Hera or Aphrodite and so on' which is different from 'I worship Zeus or I worship Hera or. . .' If there is no such sense, as would seem to be the case, this speaks against the analysis. It certainly does not refute it, though. We might just suppose that verbs such as 'worship' do not possess the power of confusing the supposition of terms following them, despite their ability to ampliate; but this appears somewhat ad hoc.

3.7.3 Ockham on Substitutivity

Let us now turn to Ockham's view concerning the problem of substitutivity. Ockham simply accepts substitutivity in non-propositional intentional contexts. Thus, in his discussion of *De Sophisticis Elenchis*, he insists that there is no fallacy in the argument:

(6)
You know Coriscus.
Coriscus is the hooded man.
———
You know the hooded man.

The inference is valid, he says.[36] His explanation of why it appears to fail is that there are similar arguments that are fallacious. The arguments in question are fallacies of accident.

What is a fallacy of accident? The term was coined by Aristotle in ch. 24 of *De Sophisticis Elenchis*, but his comments are difficult to decipher,[37] and different medieval commentators fastened onto different aspects of his

[36] Ockham (1979), 231. Here and in subsequent examples, his actual predicate is 'coming', not 'hooded'.

[37] What he actually writes about the Hooded Man is as follows ($179^{a}33$–$^{b}3$). 'Do you know the hooded man? . . . in the case of a man wearing a hood, ["to be hooded"] is not the same thing as "to be Coriscus". So suppose I know Coriscus, but do not know [the hooded man], it still isn't the case that I both know and do not know the same man.'

discussion.[38] But ‘accident’, here, it should be noted, has nothing much to do with the usual notion of accident in Aristotle. Peter of Spain wrote (about one hundred years before Ockham and Buridan):[39]

It must be said that . . . ‘accident’ is not used as it is by Porphyry as one of the five predicables [species, genus, differentia, property, and accident], nor as Aristotle uses it of the four predicates in the *Topics* [definition, property, genus, and accident], nor in the sense of accident contrasted with substance . . . But accident here means ‘does not follow of necessity’.

In the same vein, Ockham writes in his *Summa Logicae*:[40]

On this matter it should be realized that ‘accident’ is not here taken in the way it was taken earlier, where it was shown that accident is one of the five universals, but here ‘accident’ is taken for every term which can be the subject or predicate distinct from another. Whence every term which can be the subject or predicate of a proposition can be, and is, the accident in respect of another, because it is capable of being a predicate or subject distinct from another predicable.

A fallacy of accident occurs, according to Ockham, whenever one confuses an invalid syllogism with a valid one. Consequently, he says, we cannot give a general rule to describe fallacies of accident, since there are many ways of doing this. None the less, he proceeds to this much of a generalization: one type of fallacy of accident[41] occurs when a mode (such as ‘know’ or ‘possible’) is prefixed to one premise in a valid syllogism, but cannot validly be prefixed to the conclusion even though the other premise is true.[42] Thus, the syllogism:

(7)

Coriscus is a man.
Coriscus is the hooded one.
———
The hooded one is a man.

is valid. But the result of prefixing ‘You know that’ to the first premise and the conclusion is not. (None the less, if we prefix ‘You know that’ to

[38] e.g. some latched onto the thought that there is lack of unity in the middle term of the offending syllogism—suggesting a fallacy of four terms; others that the middle term does not apply in the same respect as the major term applies—suggesting a fallacy of reduplication (*S* is *M*, but not *qua P*).

[39] Peter of Spain (1972), 146.

[40] Ockham (1974), 818.

[41] In Ockham (1979) this is the second of three types of fallacy of accident; in Ockham (1974) it falls under the first of two.

[42] Ockham (1979), 239.

both premises and the conclusion, we again obtain a valid argument, he says.[43]) To take it to be valid would be a fallacy of accident.[44]

Now, at last, to come to (6). According to Ockham,[45] we take this to be invalid, since we confuse it with:

(8)
You know that Coriscus is a man.
Coriscus is the hooded one.
———
You know that the hooded one is a man.

This inference is invalid,[46] and is a fallacy of accident, since we confuse it with the valid (7).

What to make of Ockham's view might be moot, but this much is clear: there is an obvious similarity between Ockham's view and that given in the chapter. Specifically, both accept the validity of substitution in the case of intentional predicates—as in (6); and both reject it within intentional operators—as in (8).

3.7.4 Buridan on Indeterminacy and Substitutivity

We now come to Buridan on these two matters. What he has to say depends on another property of terms, *appellation*. The concept of appellation went through several phases in medieval logic. For Buridan, the appellation of a term is the concept that the term signifies, or *ratio* as he often calls it. It is therefore no different from its signification for Buridan. Where the novelty in Buridan's account lies, is in how this notion functions in intentional contexts.

According to Buridan, the appellation of an intentional term serves to restrict its range of supposition. (Restriction, as we noted, is the converse of ampliation.) Thus, consider the sentence 'You know Coriscus'—or, more accurately, 'You are one who knows Coriscus.' In this, 'one who knows Coriscus' has determinate supposition, since we can descend to 'You are person *a*, or you are person *b*, . . .', where this is an enumeration of the people who know Coriscus. But since the verb is an intentional one, its supposition range is restricted to just those people

[43] Ibid.

[44] In fact, that is perhaps a slightly misleading way to state Ockham's analysis of the fallacy of accident here. There is really only one argument, (7). What is mistaken is to suppose that one knows the conclusion of (7) if one knows its first premise (and not its second). As one might put it (clearly truly), knowledge is not closed under material consequence.

[45] Ibid. 234.

[46] According to Ockham, other commentators had misstated the paralogism.

who know Coriscus *under that appellation*. Thus, with the hooded man, one might know him under the appellation 'hooded man', but not under the appellation 'Coriscus'.[47] Appellation, note, affects the supposition only of the predicate: it has no effect on that of the subject.[48]

We can now see how this idea applies to indeterminacy.[49] Consider (1) again. In canonical form, this is 'I am one who promised you a penny.' The predicate 'one who promised you a penny' has determinate supposition, and supposits for all those who have promised you a penny—under that appellation. What could a different appellation be in this context? The identity criteria for appellations are less than obvious, but co-entailment would seem to be at least a necessary condition. Now suppose that only one person has promised you a penny, and that is me, then 'one who is 6 ft. 4 in. and promised you a penny' would be a different appellation. This term has the same supposition (namely me), but 'one who has promised you a penny' and 'one who is 6 ft. 4 in. and promised you a penny' express different concepts.

This application of the notion of appellation blocks the argument to the falsity of (1). Thus, let p_1, p_2, . . . be an enumeration of pennies. It is true that I do not owe you p_1 and I do not owe you p_2, etc. But it is not possible to ascend to 'I do not owe you a penny (p_1 or p_2 or . . .)', since the appellations of 'one who owes you a penny' and 'one who owes you p_1', etc., are different. Or, to look at in another way, the problem of indeterminacy is solved, since the denotation (supposition) of 'a penny' in (1) is no longer entering into the truth conditions of the sentence; so its denotation (supposition), whatever (or however) that is, is no longer relevant. One might still, of course, ask how its supposition enters into the supposition of the predicate 'one who offers you a penny'. But questions of this kind, concerning compositionality within terms, were not ones that the medieval logicians tended to asked themselves.

[47] See e.g. Buridan (1976), 101: 'Therefore, such verbs . . . restrict terms following them which they govern to supposit for those for which they supposit not absolutely but with the appellation of the *ratio* or the concept according to which those terms signify what they signify.' Buridan speaks here of the restriction of supposition of terms following the intentional verb. However, in applying his account of supposition, we have reinterpreted his remarks to apply to the whole predicate, including the intentional verb, in accordance with his much-repeated injunction to apply supposition strictly to the whole predicate, not to its parts.

[48] To be more precise, a subject terms supposits for something as long as it supposits under *some* appellation (Buridan (2001), 895). Effectively, then, appellation drops out of the picture.

[49] See Klima (1991).

Turning to the problem about substitutivity, it is not difficult to see how the application of appellation solves this problem too. The inference:

> Coriscus you know.
> Coriscus is the hooded one.
> ———
> The hooded one you know.

is valid, since the substitution is in the subject place, and the appellation connected with intentional verb does not restrict the supposition of this. But the inference:

> You know Coriscus.
> Coriscus is the hooded one.
> ———
> You know the hooded one.

is invalid, since the substitution is in the predicate place, and the appellation of the predicate gets in the way. As is clear, you may be one that knows a person (Coriscus—that is, the hooded one) under the appellation 'Coriscus', but not under the appellation 'hooded one'—and vice versa.[50]

According to Buridan's account, from 'You know Coriscus' one can infer 'Coriscus you know'—since the restriction of supposition generated by the intentional verb does not act on what comes before it. Thus, if Coriscus is known under that appellation, he is known (under some appellation). Thence, by substitution on the subject place, one can infer 'The hooded one you know.' But one cannot infer 'You know the hooded one.' This last move fails because of appellation. 'The hooded one you know' is true without regard to appellation, since the term 'the hooded one' occurs in front of, or before, the intentional verb 'know'. But for 'You know the hooded one' to be true, the appellation condition would need to be satisfied, that is, you would have to know him *as* the hooded one.[51]

How successful Buridan's account is, it is difficult to judge, since the full details were never worked out; but there are serious worries about it. As Ashworth writes:[52] 'to appeal to appellation theory is to acknowledge that no purely extensionalist interpretation of all propositions can be given and that no unified theory of inference is possible'. To see the problem,

[50] Buridan (2001), 896.

[51] Buridan holds the rather odd view that 'I owe you a penny' entails not just 'A penny I owe you,' but 'Every penny I owe you.' (See Klima 1991.) One cannot infer the clearly false 'I owe you every penny,' for the same reason.

[52] Ashworth (1977), 77.

just consider the sentence:

(9) I promise you every penny.

It seems pretty clear that one ought to be able to descend from this to 'I promise you this penny and I promise you that penny and. . .'. But 'promiser of every penny' and 'promiser of this penny' clearly have different appellations. Hence from the fact that 'promiser of every penny' supposits for me, it would not seem to follow that 'promiser of this penny' supposits for me. Thus, either a suppositional account of the validity of inference, in general, is impossible, or one must deny the validity of this inference.

The first line is a hard one to take. The notion of supposition can be thought of as giving, as we might put it in modern jargon, the truth conditions of sentences, and so explain the inferential relations between them. The notion of descent was central to these enterprises. Yet it is clear that the behaviour of appellation is going to get in the way of descent to singulars in cases of this kind. This would seem to rule out the term's having any suppositional mode. Some medievals operated with a notion of immobile confused and distributive supposition (though there is no evidence to suppose that Buridan did).[53] But the notion is barely coherent, since confused and distributive supposition is defined in terms of the possibility of descent, while 'immobile' means that descent is not possible for some reason.[54] In any case, without a related notion of ascent and descent, the mode of supposition is simply a name without a function.

It would seem, then, that Burdian would probably have taken the second line. In his discussion of the tenth sophism in the chapter on Appellation in his *Sophismata*[55] he states explicitly that one cannot descend from:

(10) You know every pair [of objects] to be even.

[53] Paul of Venice does. He says, for example, concerning the sentence 'You lack (a loaf of) bread' (Maierù 1972: 243, citing Paul's *Quadratura* I. 23): 'It is clear regarding the verb "lack", because it distributes and immobilizes at the same time. For from "You lack (a loaf of) bread", "You do not have (a loaf of) bread" follows, but this inference would not be valid unless "bread" in the premise stood distributively, and so "bread" stands distributively in the conclusion. But that it stands immobilely is clear, because from "You lack (a loaf of) bread" and "These are all the loaves", "You lack this loaf and that loaf and so on" does not follow.'

[54] See Paul of Venice (1971), 103. Cf. Hughes's note in Paul of Venice (1990), 230.

[55] See Buridan (2001), 893, 904.

to 'you know this pair to be even and you know that pair to be even and so on', because the appellation has changed—if you have two coins in your pocket, it does not follow that you know that the coins in your pocket are even. (You might not know how many coins you have in your pocket.) But even if Buridan is right about this particular example, the descent concerns a quantifier phrase within an intentional operator, not a predicate.[56] The inference from 'I owe you every penny' to 'I owe you this penny' is within an intentional predicate, and seems immaculate.

The basic problem here is, in fact, one familiar from modern discussions of intentionality. As is likely to be evident to the reader, Buridan's application of the notion of appellation in intentional contexts is very like Frege's application of his notion of sense in intentional contexts. Both deploy a factor other than reference (supposition) to block substitutivity in such contexts. But the cost of this is that it interferes with things such as quantification into (and other devices that bind variables within) such contexts. Frege's theory is quite different from the purely referential theories of intentionality familiar from Russell and Kripke, and implicit in the noneist account of this book. In the same way, Buridan's appeal to appellation is more at home in a non-referential account of intentionality, whilst the orthodox suppositional account is purely referential in spirit.

[56] For just this reason, the admission is not definitive; Buridan could have said that terms behave differently within intentional operators and predicates. Actually, propositional complements such as (10) (though not (9)) may cause similar problems for Ockham too. Rightly or wrongly, even he thinks that one cannot descend to a conjunction of singulars in such contexts. One can know that every truth is true without knowing that some particular truth is true. (Ockham 1979: 238.)

4 Characterization and Descriptions

4.1 Introduction: *Sein* and *Sosein*

Meinong distinguished between the *Sein* of an object and its *Sosein*. The *Sein* of an object is its existential status—which may be *none*. The *Sosein* of an object comprises the properties that it has. And Meinong insisted that an object's *Sosein* is independent of its *Sein*. That is, the existential status of an object and what properties it has are quite separate issues. The previous chapter does not endorse this claim in full generality. It insists that a non-existent object cannot have existence-entailing properties, like standard extensional ones—at the actual world, anyway. But what about its properties in general, and at worlds other than the actual? That is the topic of this chapter.

Meinongians have always insisted that if an object is characterized in a certain way, it has its characterizing properties—at least for properties of certain kinds. This is a deeply problematic claim, for reasons that are not difficult to see, and to which we will come shortly. What truth there is in this claim is the central concern of this chapter. It will turn out that objects do have the properties that they are characterized as having, provided that one understands this claim appropriately.

Descriptions ('a/the object with such and such properties') are noun-phrases that (appear to) refer to objects, and that wear their characterizations on their faces. It is therefore to be expected that the issue of characterization is closely connected with an account of the way that descriptions function. This will be our concern in the second part of the chapter.

4.2 The Characterization Principle

When we represent an object to ourselves we may do so in terms of various of its properties. Thus, we represent Holmes as living in Baker St, being a detective of acute powers of observation and inference, etc.; we, or the Ancient Greeks, represent Zeus as being the head of the Greek pantheon, as living on Mt Olympus, etc.; we, or the nineteenth-century astronomers who proposed its existence, represent Vulcan—the planet, not the god—as being a planet that has a sub-Mercurial orbit, and whose existence accounts for the precession of Mercury's perihelion; and so on. It would seem that these objects must, in some sense, have the properties that they are characterized as having. If they didn't, we wouldn't know what we were talking about when we talk about them. Moreover, we would seem to be able to think about, imagine, tell a story about, an object with any old bunch of properties that we care to put together. Thus, if $A(x)$ is any property, or conjunction of properties, we can characterize an object, c_A, and be guaranteed that $A(c_A)$. This is the *Characterization Principle* (CP) in its most naive form.

The CP cannot be accepted in this form, for it entails the existence of something satisfying any condition. Let $A(x)$ be any property. Let B be $A(x) \wedge Ex$. Applying the CP to B we get an object c_B such that $A(c_B) \wedge Ec_B$. So $\mathfrak{S}x(A(x) \wedge Ex)$. Worse, let A be any sentence one likes, and let B be $x = x \wedge A$. Apply the CP to B, and we get an object, c_B, such that $c_B = c_B \wedge A$, from which A follows. So we have proved an arbitrary A.

For this reason, no noneist has even accepted the CP in its pristine form. The standard response, from Meinong onwards, has been to accept it only if the properties deployed in the CP are of a certain kind: *assumptible*, *characterizing*, *nuclear*, the names vary. And existence (among others) is not such a predicate. The problem for this line is to give a principled characterization of what constitutes a characterizing predicate and why. No one, as far as I am aware, has been able to do this. Certain classes of predicates can be circumscribed and deemed safe. But without an appropriate rationale, it is difficult to avoid the feeling that the class has been gerrymandered simply to avoid problems. The situation is, in fact, worse than this. As observed, it would appear to be the case that we can think about an object satisfying any set of conditions whatsoever. Phenomenologically, at least, there is absolutely no difference between contemplating an object that has only officially characterizing properties—whatever those are—and one that has some non-characterizing properties as well, say existence.

Qua object of thought, each object seems to have all the properties deployed. Drawing distinctions within these properties seems entirely unmotivated. More: let *a* be represented as an existing evil demon, and *b* be represented as a purely fictional evil demon. A person may well fear *a*, but not *b*, precisely because they take the first to exist, but not the second. Even 'non-characterizing' predicates, such as existence, must, therefore, be relevant to the identity of an object. So what are we to say about the matter?

Qua object of thought, I said, an object characterized in a certain way has all the properties deployed in its characterization. This suggests an answer. Let $A(x)$ be *any* condition; this characterizes an object, c_A. And $A(c_A)$ *is* true—maybe not at this world, but at other worlds. Which? Cognitive agents represent the world to themselves in certain ways. These may not, in fact, be accurate representations of this world, but they may, none the less, be accurate representations of a different world. For example, if I imagine Sherlock Holmes, I represent the situation much as Victorian London (so, in particular, for example, there are no aeroplanes), but where there is a detective that lives in Baker St, and so on. The way I represent the world to be is not an accurate representation of our world, but our world could have been like that; there is a world that *is* like that. More precisely, there are many such worlds, since the representation is incomplete with respect to many details, e.g. whether the detective was left-handed or right-handed. Similarly, when the nineteenth-century scientists postulated the existence of Vulcan, they represented the situation as one in which there was a sub-Mercurial planet whose existence caused the precession of Mercury's perihelion—but they represented it as being governed by Newtonian dynamics, not those of Special Relativity. Now the world is, as it turns out, not like that, but there certainly are worlds where things are like this—many different ones, since the representation is incomplete in many ways.

Now, I suggest, the object characterized by a representation has the characterizing properties, not necessarily in the actual world, but in the worlds (partially) described by the relevant representation.[1] Thus, Holmes has the properties he is characterized as having not at this world, but at those worlds that realize the way I represent the world to be when I read the Holmes stories. And Vulcan has the properties it is characterized as having at those worlds that realize the theory of the nineteenth-century scientists who postulated its existence. Thus, let Φ be an intentional

[1] Similar ideas were put forward by Griffin (1998) and Nolan (1998).

operator of the form '... represents ... as holding [in the matter at hand]'. Note that each matter at hand will occasion a different predicate: if I read a novel I represent the world of the novel to myself in a certain way; if I then go to the lab and work on the Vulcan theory the act of representation is quite different. Let $A(x)$ be any condition; someone can intend an object of thought characterized by $A(x)$, and let 'c_A' rigidly designate it.[2] Then we may not have $@ \Vdash^+ A(c_A)$, but if a is the relevant agent, and Φ is the appropriate intentional operator, we do have $@ \Vdash^+ a\Phi A(c_A)$; so at every w such that $@R_\Phi^{\delta(a)}w$, $w \Vdash^+ A(c_A)$. In particular, then, if B is the condition $x = x \wedge A$, $c_B = c_B \wedge A$ is true in a certain set of worlds that may not include the actual.[3] In this way, the CP can be accepted *in full generality*: we just do not assume that an object characterized in a certain way has its characterizing properties at the actual world, only at the worlds which realize the way the agent represents things to be in the case at hand. Notice that, though it is a priori that a characterized object has its characterizing properties in the appropriate worlds, claims of the form $A(c_A)$ may certainly not be necessarily true. There is no reason to hold that they are true in all possible worlds, or even in the actual world.

4.3 Further Comments

Let me make a number of other comments relevant to this account. First, the operator Φ has a feature not shared by all intentional operators. Representations must have a minimal coherence. In particular, they must be closed under some notion of logical consequence. One way to see this fact is to note that we argue about how things are in the relevant situations; in the process, we draw inferences. Thus, we may read a Holmes novel and then argue about where Holmes was at a certain time, not made explicit by Doyle. We infer that he could not have been in Scotland because he was in London the next day (and there are no plane flights). The scientists who postulated Vulcan inferred various of its effects from its position and weight. And so on. Thus, we have:

$$\text{if } @R_\Phi^d A \text{ for all } A \in \mathcal{S}, \text{ and } \mathcal{S} \vdash_L B \text{ then } @R_\Phi^d B$$

[2] Thereafter, and assuming that something like the causal theory of names is right, others may refer to it by the same name. For further discussion of reference, see 7.5.

[3] Similarly, let B be $x = x \wedge \Box A$. Then by the CP, $c_B = c_B \wedge \Box A$ is true at some world, w. It does not follow that A is true at this world, however, since w may not be a possible world. Indeed, if $\Box A$ is not (actually) a necessary truth, it is not.

Here, we may take $\vdash_L$ to be specified by some set of rules of proof. It follows from this condition that:

$$\text{if } @ \Vdash_s^+ a\Phi A \text{ for all } A \in \mathcal{S}, \text{ and } \mathcal{S} \vdash_L B \text{ then } @ \Vdash_s^+ a\Phi B$$

In particular, then, an object characterized in a certain way has not only its characterizing properties at the appropriate worlds, but those that follow from them as well.

What notion of logical consequence is in question here? The default assumption is that it is the true notion of validity—the logic determined by the semantics I have given, assuming them to be right.[4] But as we shall see in Ch. 6, in unusual contexts, a different notion of consequence may be appropriate.

Note that, though Φ is closed under logical consequence, there may be many other intentional states associated with the situation in question that are not closed in this way. For example, suppose that I fear something (real or imagined). Then I represent the situation to myself in a certain way. But I may not fear all the consequences of what I fear, simply because I may not realize them. In particular, the fact that I represent things to be in such and such a way (Φ) does not imply that I realize that I do. Realization is certainly not closed under logical consequence. Thus, the representation may have objective features that go beyond my recognition of them.

Secondly, if the CP is to hold in full generality in the way I have explained, then for any characterization, $A(x)$, there must be worlds in which this characterization is satisfiable. In particular, there must be inconsistent worlds, since we can consider inconsistent characterizations. For example, let $A(x)$ be the property of being round and not round, $Rx \wedge \neg Rx$. It is easy to construct an interpretation where $@ \Vdash^+ a\Phi(Rc_A \wedge \neg Rc_A)$. Quite generally, it is a feature of the semantics of Ch. 1 that any way of describing things is realized at some world. Thus, if $\mathcal{S}$ is any set of sentences, some of which may have the variable x free, we may simply take an open world, w, where for every $\sigma(x) \in \mathcal{S}$, $\delta(\sigma(c), w) = \{\langle\rangle\}$ (and if $\sigma(c) \notin \mathcal{S}$, then $\delta(\sigma(c), w) = \phi$). Note that $\mathcal{S}$ could be closed under deducibility, making w a world that may realize a notion of representation, Φ.

4 In this book, I am not concerned with questions of axiomatizability, completeness, and so on. I am sure that the semantics that I describe give a logic that has a sound and complete proof theory, though I have not worked through the details. If this were wrong, we could simply let $\vdash_L$ be the notion of proof for a suitably strong fragment of it.

Third: if I represent things to myself in a certain way, the representation may not characterize the actual world, but it may. The scientists who postulated the existence of Vulcan could, after all, have got it right. The scientists who postulated the existence of the planet Uranus, and who characterized it in such and such a way did get it right. Thus, it is quite possible that for particular acts of representation, Φ, we have $@R^d_\Phi@$.

A natural question at this point is that of when characterization succeeds in picking out an existent object. Consider Holmes. He has his characterizing properties in the worlds that Conan Doyle imagined. But the actual world is not among these. Holmes exists in those worlds, but not at this one. But now consider Vulcan. As a matter of fact, the worlds that are as represented by the Vulcan-theory do not include the actual. Just as for Holmes, Vulcan is therefore an object that does not exist at this world. But suppose that $@$ had been the way that the Vulcan-theory has it, that there were a sub-Mercurial planet which, in fact, caused the precession of Mercury's perihelion. Then 'Vulcan' must refer to that. For $@$ itself is one of the worlds that $@$ accesses under the appropriate R_Φ. Hence, Vulcan satisfies its characterizing conditions at $@$. But, at $@$, the planet is the unique object satisfying these conditions. Hence, this is Vulcan. This line of reasoning occasions a certain possible objection, but I will defer discussion of it until a more general discussion of fictional objects in Ch. 6.

4.4 Identity

Let us turn now to the question of identity. Many (such as Quine, as we will see in the next chapter) have felt that there is something odd about identity when it comes to non-existent objects. When are two such objects the same: or better, when do two names for non-existent objects refer to the same object? A simple answer may be given to this question, however. Any objects, d and e, are the same iff they have the same (actual) identity. That is, at $@$, $d = e$. If this is the case, then, for all $w \in \mathcal{C}$, s and $A(x)$, where x is not in the scope of an intentional propositional operator:[5]

$$w \Vdash^+_{s(x/d)} A(x) \text{ iff } w \Vdash^+_{s(x/e)} A(x)$$

[5] See Lemma 8 of 2.12.

We may take the converse, in essence, to be our criterion of identity. This answer, note, is perfectly general, and applies whether or not the objects in question exist at @—or, for that matter, at any world.

The criterion amounts to saying that d and e have the same atomic properties at each world in $\mathcal{C}$. (The rest follows by induction.) We need to exempt identity itself from these properties, or the criterion becomes trivial. (Given that $@ \Vdash^+_{s(x/d,y/d)} x = y$ iff $@ \Vdash^+_{s(x/e,y/d)} x = y$, and that the left-hand side is true, it follows that the right-hand side is true; that is, that $d = e$ (at @).) In what follows in this section, then, talk of properties is to be understood as excluding identity. It might be thought that one should also exclude intentional properties from the criterion: two objects cannot be different objects simply because, say, I am thinking about one and not the other. I can do this only because the objects are distinct in the first place. In fact, I think this is wrong: two objects might be exactly the same, except that I am thinking about one, but not the other: that is the only thing that distinguishes them. However, it does no harm to add this further restriction if one wishes. Whatever the precise understanding of the condition, the identity of an object supervenes on the properties it has.

To apply the criterion in some cases, we need more information about the properties of characterized objects. We saw that such an object has its characterizing properties (plus all those that follow from these according to the appropriate notion of deducibility) at all the worlds that realize the way the agent represents things to be. What other properties does it have? Consider a characterized object, such as Holmes. It is natural to think that there is no more to Holmes than Doyle tells us about him: characterized objects have *only* those properties determined by the appropriate representation—call these *determined* properties—there is nothing more to them than they are represented as being. This cannot be right, though. In a closed world that realizes the Holmes stories, Holmes is either left-handed or right-handed (or ambidextrous).[6] So in every such world Holmes is left-handed or he is right-handed (or ambidextrous). At each such world he has one or other of these properties, though neither is determined.

In the light of this kind of example, it may be suggested that what the intuition really amounts to is that if a property is not determined,

[6] Let w be that world. Provided that the logic under which the representation is closed contains *modus ponens*, w is closed. For if $w \Vdash^+_s A$ and $w \Vdash^+_s A \rightarrow B$, then $w \Vdash^+_s B$.

it can vary arbitrarily across the worlds that realize the representation. Even this is not right though. Recall that a characterized object may well exist. Suppose that c_A is a characterized object, and that it is, as a matter of fact, Uranus. Then in every closed world, and a fortiori every closed world that realizes the way that things are represented as being, if Uranus has some property, so does c_A, even though this may not be determined. (Identity does not vary across closed worlds, recall.) It might be suggested that the principle applies only to non-existent objects. But even this cannot be right. Suppose that we characterize an object by the predicate Tx, 'I am thinking of x.' Then in all worlds that realize the appropriate representation, Tc_T. Suppose that the actual world is such a world, and that, as a matter of fact, I am thinking of Holmes and only Holmes. Then, as with the Vulcan example, Holmes $= c_T$. So in every closed world that realizes the representation, if Holmes has some property, so does c_T, even though this may not be entailed by the representation relevant to c_T.

What the intuition in question is, in fact, tracking is not that characterized objects can vary arbitrarily outside their determined properties, but that they are free to vary subject only to the constraints imposed by existing objects (such as myself and Uranus). This is the degree of freedom that they have. Let us call this the *Principle of Freedom*: given a characterized object, for any property that is not determined, there will be closed worlds, realizing the representation in question, in which the object has that property and ones in which it does not, subject only to constraints imposed by facts about objects that actually exist.

Let me now illustrate the criterion of identity by applying it in an example. Consider Holmes, as characterized in the representation of the Holmes stories by Doyle. Is Holmes George Bush? No. George Bush exists at this world (alas), but Holmes does not. They do not share all their properties at some closed worlds, in particular this one. Hence, they are not the same. But what of Pegasus, as represented by the Ancient Greeks? Neither Holmes (h) nor Pegasus (p) actually exists. Are they identical? Well, consider a closed world that realizes the Holmes stories. In all such worlds, Holmes is a detective, Dh. Does the representation entail that both Holmes and Pegasus are detectives, $Dh \wedge Dp$? Clearly not. Nor is there any fact about existent objects (of the sort just discussed) that determines otherwise. So by the Principle of Freedom, there will be closed worlds of this kind where $Dh \wedge Dp$ fails, and so in which Dp fails. By the criterion of identity, Holmes and Pegasus are distinct.

Whilst we are on the subject of identity, it is sometimes thought that world-semantics gives rise to extra problems concerning the notion of trans-world identity. What makes an object, x, at one world, the same object as one, y, at another? But in the semantics we have, objects are just objects; they are not 'at one world' or 'at another'. They have various properties at different worlds, but they are just themselves. (Thus, technically, they are not world-bound entities, but functions from worlds to identities.) But, it may be replied, they may have different identities (i.e. members of Q) at different worlds. What makes the different identities pertain to the same object? One can, in fact, ask exactly the same question about any aspect of an object. A (concrete) object has different colours (or heights, or weights) at different worlds. What makes the different colours (or heights, or weights) pertain to the same object? Well, that's just the way things are at that world: at that world, that object has that colour (or height, or weight). Same for identity. That an object has a certain identity is just how things are at that world. Thus, in the Hooded Man example of 2.8, there are open worlds (accessible under the accessibility relation for the appropriate intentional operator) where Nescio shares his identity with Cain, and worlds where he doesn't.

Sometimes, supposed problems concerning trans-world identity are posed employing symmetric situations. Thus, let x_1 and x_2 be objects such that, at w_1, x_1 has all the properties in $\mathcal{S}_1$, and x_2 has all the properties in $\mathcal{S}_2$; but at w_2, x_1 has all the properties in $\mathcal{S}_2$ and x_2 has all the properties in $\mathcal{S}_1$. Why, at w_2, is it x_2 that has the properties $\mathcal{S}_1$, and not x_1?[7] But why shouldn't it be? The identity of an object is not determined by its properties at any one world. As Kripke notes,[8] the problem seems to arise because, when one asks a question of this kind, one is thinking of oneself as viewing w_2 (through a sort of trans-world telescope) and having to figure out which object at that world is x_1 (and which is x_2). But identity is not determined by features intrinsic to a world, and so cannot be 'figured out' in that way.

One can, in fact, strengthen the symmetry considerations. Take any two objects, x_1 and x_2. Let w_1 be *any* world, where x_1 has the properties in $\mathcal{S}_1$ and x_2 has the properties in $\mathcal{S}_2$. Then there is a world, w_2, where x_1 has the properties $\mathcal{S}_2$ and x_2 has the properties $\mathcal{S}_1$. (And, of course, vice versa.) What, one might ask, makes x_1 x_1 and not x_2? The answer is

[7] See e.g. Chisholm (1967). [8] Kripke (1977), 82.

as before: x_1 is the object it is because it has the properties it has at the worlds where it has them, and not at others. The symmetric are always with us. Thus, there could be a *single* world in which there was nothing but two perfect spheres, x_1 and x_2, of exactly the same size, constitution, etc., standing a metre apart (and, we suppose, at least for the sake of illustration, that spatial properties are relational). Then for any property of x_1 (maybe involving x_2), there is a symmetric property of x_2 (maybe involving x_1). But x_1 and x_2 are distinct for all that.

In a similar way, in the arithmetic of complex numbers, $+i$ and $-i$ have symmetric properties. It would make no mathematical difference if what we now call $+i$, we called $-i$, and vice versa.[9] They are two none the less, as required by the fact that the equation $x^2 = -1$ has two roots.

4.5 Indefinite Descriptions

We now turn to the subject of descriptions. A description of an object packs a certain characterization into its very syntax. The semantics of descriptions might therefore be expected to be closely connected to the issue of characterization. And so it is.

Let us start with indefinite descriptions; we will come to definite descriptions in the next section. Indefinite descriptions are of the form 'a(n)/one (particular) object, x, such that $A(x)$'. I will write this as $\varepsilon x A(x)$. Formally, we extend the language of previous chapters with the operator ε. If $A(x)$ is any formula, $\varepsilon x A(x)$ is a term of the language. Thus, terms and formulas are now defined by a joint recursion, since predicates turn terms, including descriptions, into formulas, and descriptions turn formulas into terms.[10]

Denotation and truth-value are defined, similarly, by a joint recursion. Truth/falsity conditions are exactly as before. The denotation conditions for terms other than descriptions are also the same as before. We therefore need concern ourselves only with the denotation conditions for descriptions.

To give the semantics for descriptions, a new component, φ, has to be added to interpretations.[11] Thus, an interpretation is now of the form $\langle \mathcal{P}, \mathcal{I}, \mathcal{O}, @, D, \delta, \varphi \rangle$. φ specifies a collection of choice functions.

[9] Technically, the function that maps each complex number to its conjugate is an automorphism of the complex plane.

[10] See e.g. Leisenring (1969), ch. 1.

[11] The account that follows derives largely from Priest (1979).

Specifically, if τ is any descriptive term, $\varphi(\tau)$—or φ_τ as I will write it—is a function from subsets of D to D, such that if $X \subseteq D$ and $X \neq \phi$, $\varphi_\tau(X) \in X$. The machinery is a bit more complex than that usually deployed in semantics for descriptions: it would be more normal to use just a single choice function. I will discuss the reason for the present approach in the next section. We will also need the notion of the matrix of a term. This is defined in the same way as the matrix of a formula (see 1.5). Thus, an ε-term matrix is any ε-term in which all the free terms are distinct variables, and these are the least ones greater than all the bound variables in the term (in some canonical ordering), in increasing order from left to right. As with formulas, every ε-term, τ, has a unique matrix, $\overline{\tau}$, from which it may be obtained by the substitution of terms for the variables.

Let τ be the description $\varepsilon x A(x)$. Its denotation is one of the things satisfying $A(x)$ if something does; otherwise it is some other object. More precisely, $\delta_s(\tau)$ is:

$\varphi_{\overline{\tau}}\{d : @ \Vdash^{+}_{s(x/d)} A(x)\}$	if this set is non-empty
$\varphi_{\overline{\tau}}(D)$	otherwise

The reason that $\overline{\tau}$ is used, rather than τ, is that it makes the denotation of τ independent of which particular free terms are employed in τ. It is only their denotations that are relevant. This allows quantification to work properly within the scope of descriptive terms, as the technical appendix to this chapter demonstrates.

Now let $d = \delta_s(\tau)$. If anything has its defining characteristics, it does. No constraints have yet been put on d in the other case (not even that it does not exist). But there is more to be said. Even if d does not satisfy $A(x)$, an agent, a, contemplating τ still represents d as satisfying $A(x)$. Thus for example, a person can imagine a (particular) green elephant in their room. Even if there is no such object, it remains the case that that thing is green, an elephant, and in their room—at least in their imagination. So the object d must at least be an object that satisfies $A(x)$ in those worlds that realize a's representation of the situation. That is, it should be the case that $@ \Vdash^{+}_{s(x/d)} a\Phi A(x)$, where Φ is an intentional operator of the kind deployed in 4.2.

But what is that representation? This may well depend on the circumstances, since we employ descriptions in talking about all kinds of situations. But in all of them we at least represent d as satisfying $A(x)$.

Thus we have, for any e:

$$(\star) \quad \text{if } @R^e_\Phi w \text{ then } w \Vdash^+_{s(x/d)} A(x)$$

This condition therefore imposes a constraint on an acceptable interpretation for the language, which we henceforth assume.[12] It follows, in virtue of the constraint, that $@ \Vdash^+_s a\Phi A(\varepsilon x A(x))$; thus, $\models a\Phi A(\varepsilon x A(x))$. Also, since Φ is closed under an appropriate notion of logical consequence, if B follows from $A(\varepsilon x A(x))$ according to that notion, we have $@ \Vdash^+_s a\Phi B$ as well.

Notice that the preceding treatment of descriptions makes them rigid designators. The denotation of a description is the same at all worlds. This is different from the more usual treatment of descriptions, according to which they are not necessarily rigid.[13] It is certainly possible to give an account of this kind. (The denotation of $\varepsilon x A(x)$ would then be world-dependent. Crucially, its denotation at w would be selected from $\{d : w \Vdash^+_{s(x/d)} A(x)\}$, where this is non-empty.) One may even give an account in which there are two sorts of descriptions, one rigid and the other not. But employing a rigid notion of descriptions makes life easier (in, particular, concerning quantifiers); so let us settle for simplicity on this occasion.[14]

4.6 Definite Descriptions and Speaker-Intention

Let us now turn to definite descriptions. These are things of the form 'the object, x, such that $A(x)$'. I will write this as $\iota x A(x)$. How should one handle definite descriptions? The only difference between these and indefinite descriptions is that the former, but not the latter, impute uniqueness. Thus, the simplest approach is to define $\iota x A(x)$ as $\varepsilon x(A(x) \wedge \neg\mathfrak{S}y(A(y) \wedge y \neq x))$, which we henceforth do.

Note that the semantics of definite descriptions require the use of choice operators, φ_τ, just as much as those of indefinite descriptions. If one is not a noneist then, depending on how denotation-failure is handled,

[12] Note that the operator Φ depends on the act of representation in question, and therefore on $A(x)$. Note, also, that the condition is impredicative, in the sense that $A(x)$ may itself involve the operator Φ, and hence its truth conditions may involve R_Φ.

[13] See e.g. Fitting and Mendelsohn (1998), ch. 12.

[14] Essentially, if descriptions are not rigid then what is expressed in the text as $A(\varepsilon x B(x))$ is expressed by $\mathfrak{S}y(y = \varepsilon x B(x) \wedge A(y))$.

a choice function may not be required; but a noneist requires a denotation in those cases where either nothing or more than one thing satisfies $A(x)$. Formally, this case has to be handled by a choice function.

Which brings me to the question of the significance of choice functions. The deployment of a choice function is a recognition of the fact that, as far as the formal semantics go, the denotation of the descriptive term is non-deterministic. That is, the denotation of the term is something that is determined by factors outside the semantics. Principal amongst these is context, and especially speaker-intention, similar to the way that this is relevant to determining the referent of a demonstrative (a main difference being that a person can intend pretty much anything they like with a simple demonstrative, whilst the content of a description puts constraints on what can be intended). Thus, suppose you say (truly), for example: 'I saw a man on the tram I was on yesterday; he looked rather sad.' The referent of 'a man on the tram I was on' in this context is the particular man whom you saw, and to whom you now intend to refer. (Note that there could have been more than one sad-looking man on the tram; but you are talking about a particular one of them.) Of course, you could be lying: the man on the tram was not sad. The description refers to him none the less. Maybe you didn't even get on a tram at all. In that case, the description refers to the presumably non-existent object intended in your imagination. Or suppose that I ask you to imagine three red squares. You do so; then you tell me (truly) that you are thinking of one particular one of them, employing that very indefinite description. Perhaps this is the one in the middle. Then the denotation of the description is that particular object, and it is so because of your intentional act.[15]

Why do we need multiple choice functions? Suppose that φ itself were a choice function, and that in the denotation conditions for descriptions '$\varphi_{\overline{\tau}}$' were simply replaced by 'φ'. Then extensionality for descriptions would follow. For suppose that $\{d : @ \Vdash^{+}_{s(x/d)} A(x)\} = \{d : @ \Vdash^{+}_{s(x/d)} B(x)\}$. Then whether or not $\{d : @ \Vdash^{+}_{s(x/d)} A(x)\}$ is empty, the set from which the denotations of $\varepsilon x A(x)$ and $\varepsilon x B(x)$ are selected are the same. It would follow that $\delta_s(\varepsilon x A(x)) = \delta_s(\varepsilon x B(x))$, and so, in particular, that $@ \Vdash^{+}_{s} \varepsilon x A(x) = \varepsilon x B(x)$. Not only is this an outcome that has little to be said for it intuitively; in the present context, it is certainly wrong, since it misrepresents the intentional act.

[15] Strictly, then, the choice function should also depend on the agent in question. I ignore this extra complication in the formal semantics.

Thus, consider the descriptions 'the mermaid in my room' and 'the centaur in my room'. The extensions of 'mermaid in my room' and 'centaur in my room' are both empty at the actual world, @. If we had extensionality, it would follow that it is true that the mermaid in my room is the same as the centaur in my room. But this seems wrong. I can, for example, imagine the mermaid in my room without imagining the centaur in my room.[16]

4.7 Properties of Descriptions

The preceding semantics for descriptions are very natural. Let me briefly review the properties of descriptions that they deliver. (Proofs of the facts in question can be found in the technical appendix to this chapter.) Since descriptions are rigid, they behave as one would expect with respect to quantifiers. Thus, for example, they satisfy $\mathfrak{A}$-elimination and $\mathfrak{S}$-introduction. The treatment also ensures that quantification into, and substitution of identicals into, the scope of a description behave properly (provided, in the latter case, that we are not substituting into the scope of an intentional operator).

The denotation conditions of descriptions suffice to ensure that $\mathfrak{S}xA(x) \models A(\varepsilon xA(x))$ (provided that $\varepsilon xA(x)$ is free when substituted in $A(x)$). Let us write $\mathfrak{S}x(A(x) \land \neg\mathfrak{S}y(A(y) \land y \neq x))$ as $\mathfrak{S}!xA(x)$. Then because of the definition of ι, it follows that $\mathfrak{S}!xA(x) \models A(\iota xA(x)) \land \neg\mathfrak{S}y(A(y) \land y \neq \iota xA(x))$, and hence $\mathfrak{S}!xA(x) \models A(\iota xA(x))$. Thus if something satisfies the defining condition of a description (and a fortiori, if there exists something which does so), the thing denoted by the description does so. But note that even descriptions denoting non-existent objects may satisfy their defining conditions. For example, provided that $\delta^-(E, @) \neq \phi$, $\varepsilon x\neg Ex$ denotes a non-existent object, but $@ \Vdash^+ \neg E\varepsilon x\neg Ex$.

[16] Using different choice functions for different terms individuates a term's denotation, effectively, in terms of its syntax. It might be argued that this is too strong. For example, arguably the mermaid in my room is identical with the creature in my room with the top half of a woman and the bottom half of a fish. If this is the case, then φ can be constrained appropriately. Thus, the following condition ensures that co-entailing conditions pick out the same object: if, for all s, $@ \Vdash^+_s \mathfrak{A}x(A(x) \leftrightarrow B(x))$ then $\varphi_{\varepsilon xA(x)} = \varphi_{\varepsilon xB(x)}$.

We are not guaranteed the condition $A(\varepsilon x A(x))$ in general. However, whether or not $\varepsilon x A(x)$ exists, we will always have $@ \Vdash_s^+ a\Phi A(\varepsilon x A(x))$. The object denoted by a description has its characterizing properties at least in the way the world is represented as being. It also has those properties that follow from the characterization according to the appropriate notion of deducibility.

Let me conclude this section by considering one possible objection to these semantics. Suppose that I believe there to be a man next door who is nasty and vicious. I have never seen him, though I have been told about him. I fear him. But suppose also that, in reality, though there is a man next door, he is meek and mild and a very friendly person. Now consider the sentence:

> I fear the man who lives next door.

This would seem to be true, but on these semantics, 'the man who lives next door' refers to the meek and mild man who actually lives next door. But surely I don't fear him. It is the nasty, vicious man next door that I fear. But for just that reason, the sentence in question is actually false. Though one might use it as a shorthand to express one's fears, what is actually true is:

> I fear the man who is nasty and vicious and lives next door.

And here, as one would expect, the description 'the man who is nasty and vicious and lives next door' refers to the non-existent object of my nightmares.

4.8 Conclusion

In the last two chapters, I have provided an account of the semantics of intentional predicates and their objects. This chapter, in particular, has provided an account of the way that characterization works, together with a closely related account of descriptions. The chapters in the first part of the book together provide a unified account of the semantics of a language with quantifiers, descriptions, identity, a relevant conditional, modal operators, and intentional operators and predicates. Noneism is absolutely central to these semantics. Some of the problems supposed to be associated with noneism, such as those concerning characterization, have been taken up along the way. Others remain—or at least, may be thought to do so. We will turn to these in the second part of the book, together with some of the more natural applications of noneism other than those directly concerning intentionality.

4.9 Technical Appendix

In this appendix, I will verify the various technical claims made in the chapter. We start by showing that quantifiers behave properly.

Lemma 11 *Fix any interpretation. Let t and A be any term and formula. Then if s_1 and s_2 are any evaluations of the variables that agree on the variables free in t and A:*

1. $\delta_{s_1}(t) = \delta_{s_2}(t)$
2. for all $w \in \mathcal{W}$, $w \Vdash^{\pm}_{s_1} A \Leftrightarrow w \Vdash^{\pm}_{s_2} A$

Proof The proof extends that for Lemma 5 of 2.12. Because of the interconnection between terms and formulas, we now have to prove the result by a joint recursion. For 1, the cases for constants, variables, and function symbols are the same as before. For descriptions, let τ be $\varepsilon y A(y)$. Then:

$$\delta_{s_1}(\tau) = \varphi_{\overline{\tau}}\{d : @ \Vdash^{+}_{s_1(y/d)} A(y)\} \quad \text{if this is non-empty}$$
$$\varphi_{\overline{\tau}}(D) \quad \text{otherwise}$$

The first of these sets may be replaced by the corresponding set for s_2 by 2. The result follows.

The proof of 2 is exactly as in Lemma 5. ■

Lemma 12 *Fix any interpretation. Let $t'(x)$ and $A(x)$ be any term and formula. Let t be any term that can be freely substituted for x in these. Let s be any evaluation of the free variables, then if $d = \delta_s(t)$:*

1. $\delta_{s(x/d)}(t'(x)) = \delta_s(t'(t))$
2. for all $w \in \mathcal{W}$, $w \Vdash^{\pm}_{s(x/d)} A(x) \Leftrightarrow w \Vdash^{\pm}_{s} A(t)$

Proof The proof extends that for Lemma 6 of 2.12. Because of the interconnection between terms and formulas, we now have to prove the result by a joint recursion. For 1, the cases for constants, variables, and function symbols are the same as before. For descriptions, let $t'(x)$ be

$\varepsilon yA(x)$. If y is x then x is not free in t', and the result follows from the previous lemma. If not, since t is free in $\varepsilon yA(t)$, this term and the term $\varepsilon yA(x)$ have the same matrix. Let this be τ. Then:

$$\begin{aligned} \delta_{s(x/d)}(\varepsilon yA(x)) &= \varphi_\tau\{e : @ \Vdash^+_{s(x/d,y/e)} A(x)\} && \text{if this set is empty} \\ &\quad \varphi_\tau(D) && \text{otherwise} \\ &= \varphi_\tau\{e : @ \Vdash^+_{s(y/e)} A(t)\} && \text{if this set is empty} \quad (\star) \\ &\quad \varphi_\tau(D) && \text{otherwise} \\ &= \delta_s(\varepsilon yA(t)) \end{aligned}$$

For $(\star)$, note that since t is free when substituted, it cannot contain y free. Hence, $\delta_s(t) = \delta_{s(y/e)}$ by Lemma 11, and we can apply the Induction Hypothesis in the form where s is $s(y/e)$.

The proof of 2 is exactly as in Lemma 6. ∎

The correct behaviour of the quantifiers follows as in Corollary 7 of 2.12. We next turn to identity.

Lemma 13 *Fix any interpretation, with $d, e \in D$. Let $d(@) = e(@)$. Let $t(x)$ and $A(x)$ be any term and formula such that x is not in the scope of a Ψ in $A(x)$. Then for any evaluation, s:*

1. for all $w \in \mathcal{C}$, $\delta_{s(x/d)}(t(x))(w) = \delta_{s(x/e)}(t(x))(w)$
2. for all $w \in \mathcal{C}$, $w \Vdash^\pm_{s(x/d)} A(x) \Leftrightarrow w \Vdash^\pm_{s(x/e)} A(x)$

Proof The proof is as in Lemma 8 of 2.12. The only difference is that we have to add a clause in the proof of 1 to cover descriptions. Let τ be $\varepsilon yA(x)$. If y is x then the result follows from Lemma 11. If not, then:

$$\begin{aligned} \delta_{s(x/d)}(\varepsilon yA(x)) &= \varphi_{\overline{\tau}}\{b : @ \Vdash^+_{s(x/d,y/b)} A(x)\} && \text{if this set is non-empty} \\ &\quad \varphi_{\overline{\tau}}(D) && \text{otherwise} \\ &= \varphi_{\overline{\tau}}\{b : @ \Vdash^+_{s(x/e,y/b)} A(x)\} && \text{if this is non-empty, by 2} \\ &\quad \varphi_{\overline{\tau}}(D) && \text{otherwise} \\ &= \delta_{s(e/x)}(\varepsilon yA(x)) \end{aligned}$$

The result follows. ∎

The substitutivity of identicals, except within the context of an intentional operator, follows from this Lemma, as in Corollary 9 of 2.12.

Next, descriptions behave as advertised:

Lemma 14 *Provided that $\varepsilon x A(x)$ is free in $A(\varepsilon x A(x))$:*

1. $\mathfrak{S} x A(x) \models A(\varepsilon x A(x))$
2. $\mathfrak{S}! x A(x) \models A(\iota x A(x))$

Proof For 1: suppose that $@ \Vdash_s^+ \mathfrak{S} x A(x)$. Then if $d = \delta_s(\varepsilon x A(x))$, $@ \Vdash_{s(x/d)}^+ A(x)$. By Lemma 12, $@ \Vdash_s^+ A(\varepsilon x A(x))$. For 2: suppose that $@ \Vdash_s^+ \mathfrak{S}! x A(x)$. Then by 1, $@ \Vdash_s^+ A(\iota x A(x)) \wedge \neg \mathfrak{S} y (A(y) \wedge y \neq \iota x A(x))$. Hence, $@ \Vdash_s^+ A(\iota x (A(x))$. ■

Finally, whether or not the denotation of a term is exists, we have:

Lemma 15 *Provided $\varepsilon x A(x)$ is free when substituted in $A(x)$,* $\models a \Phi A(\varepsilon x A(x))$.

Proof Choose any interpretation. Let d be $\delta_s(\varepsilon x A(x))$. Then we have that for any e:

$$
\begin{array}{rcll}
@ R_\Phi^e w & \Rightarrow & w \Vdash_{s(x/d)}^+ A(x) & \text{by condition } (*) \text{ of } 4.5 \\
 & \Rightarrow & w \Vdash_s^+ A(\varepsilon x A(x)) & \text{by Lemma 12}
\end{array}
$$

Hence, we have: for all w such that $@ R_\Phi^{\delta_s(a)} w$, $w \Vdash_s^+ A(\varepsilon x A(x))$. That is, $@ \Vdash_s^+ a \Phi A(\varepsilon x A(x))$. ■

Part II

In Defence of Non-Being

This page intentionally left blank

‘That’s bullshit.’
Richard Sylvan (in conversation).

This page intentionally left blank

5

On What There Isn't

5.1 Introduction: Quine's Critique

In the first part of the book I specified the semantics of a language with intentional operators and predicates, together with the usual logical machinery. The semantics is a noneist one. Some major objections to noneism were taken up in the process. The purpose of the second part of the book is to defend noneism against various other objections. Along the way, we will look at a couple of applications of noneism other than intentionality.

Perhaps the single most influential paper attacking noneism is Quine's 'On What there Is' (1948). The main purpose of this chapter is to look at the arguments in that. As we will see, these arguments have had an influence out of all proportion to their cogency. In fact, this was already shown by Routley in his analysis of the paper, 'On What there Isn't' (1982). It is unfortunate that this paper is relatively inaccessible, and so not better known. This fact provides the reason for this chapter, since what I say about Quine here does little more than repeat what is in Routley's paper—modulo our different accounts of characterization. The title of his paper is so good that I even decided to steal it for the title of this chapter.

Before we turn to Quine, however, it will be useful to have a look at the writer who is normally cited as first demolishing Meinong, Bertrand Russell. He and Quine make interesting foils.

5.2 Russell's Meinongianism

The first thing to note about Russell is that he himself, at one time, held a view about objects that was in some ways similar to that of Meinong.

In the *Principles of Mathematics* he wrote:[1]

> Whatever may be an object of thought, or can occur in a true proposition, or can be counted as a *one*, I call a *term* . . . Every term has being, i.e. *is* in some sense. A man, a moment, a number, a class, a relation, a chimera, or anything else that can be mentioned is sure to be a term; and to deny that such a thing is a term must always be false.

Moreover:[2]

> . . . to mention anything is to show that it is. *Existence*, on the contrary, is the prerogative of some only amongst beings. To exist is to have a specific relationship to existence—a relation, by the way, which existence itself does not have.

Though not here, Russell sometimes uses the term *subsistence* for those things that have being but do not exist, such as abstract objects—like existence.

At this time, then, Russell endorsed the view that some objects, such as a chimera, Holmes, Zeus, and so on, do not exist. But note, equally, the difference between Russell's view and that of the mature Meinong.[3] Meinong, too, thinks that some objects do not exist. He also uses the term 'subsist' (*besteht*) for some of these—essentially, those that we would call abstract objects. But some non-existent objects have no form of being at all: they neither exist nor subsist. They simply are not (they have *Nichtsein*). Thus, a chimera, Holmes, Zeus, and so on, have no being at all. Contrast this view, again, with noneism of the kind of Routley and of this book. According to this, some objects exist, or are, if you like—essentially the concrete ones in the world. All others do not (or are not). They have no being in any form whatever. Abstract objects, in particular, do not subsist. They are simply non-existent objects. On this, more in Ch. 7.

5.3 Russell's Critique of Meinong

Russell, of course, gave up his ontological view after the *Principles of Mathematics*, when he invented his theory of descriptions. He remained, however, sympathetic to Meinong's view. Thus, in a review of a collection of essays by Meinong and his students, published in the same year as 'On Denoting'—in fact, in the very same issue of *Mind* in which 'On

[1] Russell (1903), 43. [2] Ibid. 449. [3] See Meinong (1904).

Denoting' appeared—he writes:[4] 'The philosophy set forth in [the essays in the volume] is a development of that contained in Meinong's *Uber Annahmen*, and its value appears to me to be very great.' Neither is this simply politeness, as the rest of the review shows. He no longer accepts the view, of course. When he explains his objections to Meinong's view (pp. 80 f.), these target the CP in its pristine form, particularly as concerns impossible objects (which would violate the law of non-contradiction) and existence (since the CP allows us to show that anything exists). Similar arguments are deployed in the brief critique of Meinong in 'On Denoting' itself.[5] These worries are very real, as I indicated in the last chapter. But as I also showed there, there is a perfectly sensible way to assuage them.

By the time we get to Russell's *Introduction to Mathematical Philosophy*, he dismisses Meinong's view more tersely. We get:[6]

> It is argued, *e.g.* by Meinong, that we can speak about 'the golden mountain', 'the round square' and so on; we can make true propositions of which these are the subjects; hence they must have some kind of logical being, since otherwise the propositions in which they occur would be meaningless. In such theories, it seems to me, there is a failure of that feeling for reality which ought to be preserved even in the most abstract studies. Logic, I should maintain, must no more admit a unicorn than zoology can; for logic is concerned with the real world just as truly as zoology, though with its more abstract and general features.

Note that Russell has got Meinong wrong. Meinong does not claim that the objects in question have *any* form of being. What Russell is arguing against is his own earlier view! This does, indeed, offend our sense of reality. The claim that the realm of being, the totality of what is, contains chimeras and similar objects, *is* an affront to one's sensibility. We *know* that there aren't any chimeras. But this is neither Meinong's view nor that of noneism.

One could try to rework the point into an objection to noneism. One might try putting the point like this: in reality, there is no Pegasus, no Father Christmas, and no other non-existent object. But this is something the noneist agrees with! There is, i.e. exists, no such thing. Yet some objects do not exist. What someone who wishes to pursue this line of objection has to say is that it is just plain false that some objects do not exist: all objects are existent objects. But now the inadequacy of the objection is manifest: it clearly begs the question. The noneist's very claim

[4] Russell (1905*a*), 77. Page references to this and 'On Denoting' are to the Lackey reprint.
[5] Russell (1905*b*), 107–8. [6] Russell (1919), 169.

is that some objects do not exist: one can refer to them, quantify over them, etc. These objects are just not actual. The objection, then, simply collapses into a statement of actualist dogma.

5.4 On What There Is

Let us now turn to Quine's 'On What There Is'.[7] Russell was a respectful critique of Meinong; there is no respect in Quine. He never mentions Meinong by name, but it is pretty clear that he has Meinong in his sights. Certainly, many philosophers have taken Quine to have Meinong as one of the philosophers he attacks. These targets come in for a good deal of parody and ridicule, if not abuse.[8] In fact, the paper is long on rhetoric, but short on argument. (I suspect that this counts for much of the paper's influence. Rhetoric, I am sure, has much more effect on philosophers than they care to admit.)

The paper starts with two philosophical caricatures, McX and Wyman. It is not clear who McX is supposed to be; Wyman is usually taken to be Meinong. He is not. According to Wyman, all terms denote; the objects that are denoted all have being; but some of these exist (are actual) and the rest merely subsist. As is clear, this is not Meinong, even less noneism; it is Russell of the *Principles of Mathematics*. Thus when Quine says, magnanimously (p. 3): 'The only way I know of coping with this obfuscation of issues is by giving Wyman the word "exist". I'll try not to use it again; I still have "is" ' this is off-target for Meinong. Many of his objects *are* not in any sense. For the noneist, indeed, to exist and to be are *exactly* the same thing. Holmes does not exist; Holmes is not. There exists / is nothing that is Sherlock Holmes.

The next paragraph commences (p. 4): 'Wyman's overpopulated universe is in many ways unlovely. It offends the aesthetic sense of us who have a taste for desert landscapes.' Again, this may apply to Russell, but it is off-target for noneism. The non-existent objects do not overpopulate any universe, just because they do not exist, in any sense of the word. A noneist, who takes only concrete objects to exist, has a very spare universe. It is, in fact, platonists such as Quine himself who overpopulate the world with existent abstract objects that offend the aesthetic sense in question, as well as Russell's robust sense of reality. The question of the

[7] Quine (1948). Page references are to the reprint.

[8] 'Wyman, by the way, is one of those philosophers who have united in ruining the good old word "exist" ' (p. 3). 'For McX, this is an unusually penetrating speech' (p. 11).

relationship between noneism and platonism is one to which we will return in Ch. 7.

The paragraph continues with Quine's remarks about the notorious possible fat man in the doorway. Since this contains the only substantial argument that Quine has against noneism, I will come to it after this general review of the paper.

Quine then takes up the subject of impossible objects. He suggests (p. 5) that if we take contradictory descriptions such as 'the round square cupola on Berkeley College' to be meaningful, this may commit us to the contradiction that it is both round and square. But Wyman, he says, takes such phrases to be meaningless. It is not clear who Wyman is now meant to be. It is certainly not Russell, Meinong, or the noneist of this book, for all of whom the phrase is as meaningful as any other description. Nor, as we have seen in the last chapter, does it commit the noneist to a contradiction. The round square cupola on Berkeley College is, indeed, round, square, a cupola, and on Berkeley College—but not in the actual world, only in the world as we represent it to be when we think of it.

Next, Quine goes on to claim that the problems concerning non-existent objects can be solved by applying Russell's theory of definite descriptions. The procedure has two stages. First, we eliminate all proper names by replacing them with appropriate definite descriptions. We then apply Russell's theory, to eliminate all descriptions contextually.

Both phases of the strategy are problematic. Proper names, such as 'Zeus', are not equivalent in meaning to any description. Thus, 'Zeus' does not mean the same as, for example, 'the being who is the head of the Greek pantheon'. If it did, 'Zeus was the being who was head of the Greek pantheon' would be analytic, which it was not. Proper names and descriptions, in fact, hook on to reality in different ways. The arguments for this conclusion were not sufficiently appreciated when Quine wrote 'On What There Is'. They were forcefully put by Kripke (1972), and are now so well known that they hardly need rehearsing.

More importantly, even if one could eliminate all designators except definite descriptions, the second stage of Quine's strategy demonstrably fails. He considers a few examples where the strategy does seem to work; but there are others. Consider, for example, the following:

> Meinong believed that the being who is the chief god living on Olympus, lives on Olympus.

This is true. Meinong did believe that the round square was round, and so on. According to Russell's theory of definite descriptions, this sentence is ambiguous, depending on whether the scope of the description is primary or secondary. The two readings are:

1. There exists a unique being who is chief god living on Olympus, and Meinong believes that he lives on Olympus.
2. Meinong believed that there exists a unique being who is chief god living on Olympus, and he lives on Olympus.

The first of these is false, since no such being exists. But, equally, the second is false. Meinong did not believe the Greek gods to exist any more than you or I do: he knew that they are mythological. Quine, as is well known, is sceptical about quantification into intentional contexts such as that in 1. This does not help: it just makes matter worse, since it reduces the two bad choices to one.

After this (p. 9), Quine's article turns to the question of the existence of universals. We will, in fact, turn to the issue of a noneist account of abstract objects in Ch. 7. But most of the rest of the paper is of no relevance to the present topic—except for one thing. It is in this part of the paper that Quine proposes his famous thesis about existence and quantification (p. 13):

> To be assumed as an entity is, purely and simply, to be reckoned as the value of a variable.

Or, as it is often more pithily put: to be is to be the value of a bound variable. However, no real argument is given for this view. It is simply stated as a piece of dogma. It is common enough to read the particular quantifier as 'there is/exists an object such that'. Quine simply *assumes* that this is the way it must be read. Certainly, the reading follows that of Frege and Russell. But had Quine known a little about medieval logic, he would have known that this is not the only way it can be understood, as we saw in the appendix to Ch. 3.

5.5 The Possible Fat Man in the Doorway

Let us now return to the subject of the possible fat man in the doorway. I quote the passage in full (p. 4):

> Wyman's slum of possibles is a breeding ground for disorderly elements. Take, for instance, the possible fat man in that doorway; and, again, the possible bald man

in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? Are there more possible thin ones than fat ones? How many of them are alike? Or would their being alike make them one? Are no *two* possible things alike? Is this the same as saying that it is impossible for two things to be alike? Or, finally, is the concept of identity simply inapplicable to unactualized possibles? But what sense can be found in talking of entities which cannot meaningfully be said to be identical with themselves and distinct from one another? These elements are well-nigh incorrigible. By a Fregean therapy of individual concepts, some effort might be made at rehabilitation; but I feel we'd do better simply to clear Wyman's slum and be done with it.

Quine's complaint is that there is no sensible way to individuate non-existent objects. No entity without identity, in another of his pithy slogans. The very idea of such objects is therefore incoherent. But note that he makes no attempt to defend this thesis, or to show that there is no appropriate notion of identity. We simply get a string of rhetorical questions.

What is to be said about all this? First, it is not at all clear that objects require determinate conditions of individuation. Thus, there appears to be no fact of the matter as to where such objects as mountains and seas come to an end. If one walks across Australia, for example, where does one start crossing the Great Central Desert? And is New Holland (the name given by Dutch explorers to the land they found) the same place as Australia? Did the name apply, for example, to Tasmania or the Torres Straight Islands? There is no determinate answer to these questions. Yet we do not conclude that the notion of a mountain or a sea is incoherent, or that there is no such thing as Australia. Routley goes in for his own bit of parodying here:[9]

The slum of entities is a breeding ground for disorderly elements. Take, for instance, the cloud in the sky above; and, again, the adjacent cloud in the sky. Are they the same cloud or two clouds? How are we to decide? How many clouds are there in the sky? Are there more cumulus than nimbus? How many of them are alike? Or would their being alike make them one? . . . is the concept of identity simply inapplicable to clouds? But what sense can be found in talking of entities which cannot meaningfully be said to be identical with themselves and distinct from one another? These elements are well-night incorrigible . . . I feel we'd do better to clear the slum of entities and be done with it.

[9] Routley (1982), 421. Note that 'entity' is Routley's word for an existent object.

But in any case, it is quite possible to give a sensible account of the identity of objects, whether existent or not. Objects are the same iff they have exactly the same properties at all closed worlds, in the sense that I explained in 4.4. Armed with this criterion, we can now answer all Quine's rhetorical questions. In fact, it is only the first that is not entirely trivial.

- Are the possible fat man in that doorway (f) and the possible bald man in that doorway (b) the same possible man, or two possible men?

Let f be $\iota x(Fx \wedge Dx)$ and b be $\iota x(Bx \wedge Dx)$.[10] Being in that doorway—presumably Quine had in mind some actual, but empty, doorway—is an existence-entailing property. Any man in the doorway therefore exists. But there exists no one in the doorway. Hence, no one is in the doorway: in particular, f and b are not in the doorway. (We are therefore in the second case concerning the denotation of descriptions in 4.5). The question is whether f and b are one or two.[11]

There is no determinate answer to this. But not because there is indeterminacy in identity conditions. These, as we have seen, are quite determinate. There is no determinate answer because the denotations of these two descriptions depend on context, including speaker-intention, as we noted in 4.6. Thus, there is no determinate answer to the question for exactly the same reason that there is no determinate answer to the question as to whether this is that, when no context is provided. When the context is provided, a determinate answer is forthcoming. Let us look at a few examples to illustrate the point.

First, a simple example. Suppose that we are imagining a situation in which there are two particular men in the doorway. One is bald and fat; the other is bald and not fat. f and b have their characterizing properties in the worlds that realize this representation. Hence, 'f' refers to the first of these (there is no option). To which one does 'b' refer? That depends

[10] I assume that when Quine talks of a possible fat man in the doorway he does not intend the modality to be part of the definite description; he is just giving the status of the thing described. After all, suppose that I stand in the doorway. It is certainly possible that I am fat. Hence, I am a possibly fat man in the doorway. This is not the sort of thing that Quine has in mind. However, if one insists on parsing Quine's description for f as $\iota x(\Diamond Fx \wedge Dx)$, or even as $\iota x\Diamond(Fx \wedge Dx)$—and similarly for b—the matter is essentially the same.

[11] Routley's answer is 'two', since f and b have their characterizing properties, and only those entailed by them, at this world. So f is fat and b is not. This solution is not available given the treatment of characterization I have advocated.

entirely on which you intend. If you intend the first, b and f are the same; if not, they are different.

The next examples illustrate situations where representation entirely constrains intention. Suppose that we are discussing a purely imaginary situation, maybe a film that we have seen. In this, a purely imaginary character, let us call him Harry, who is fat and bald is standing in the doorway painting the frame. Since f and b have their characterizing properties in the worlds that realize this representation, 'f' and 'b' both refer to Harry: f and b are therefore one.

Next suppose that we are talking about the possibility of a fat man coming through the door. We represent this situation to ourselves. Let us suppose that we have an actual friend called Harry, who is neither fat nor bald. None the less, we imagine Harry having grown fat and coming through the doorway: 'f' denotes this man. Next, suppose we talk about the possibility of a bald man coming through the door. Now we imagine Harry having grown bald coming through the doorway: 'b' denotes this man. Both 'f' and 'b' refer to Harry. Again, they are one and the same person.

Finally, a different sort of example. Suppose that we are talking about the possibility of a fat man coming through the door. We represent this situation to ourselves; the man in question is purely imaginary: 'f' denotes that man. The representation leaves open the question of whether someone else comes through the doorway at the same time. If 'b' denotes some actual man, f and b are distinct, since one exists and the other does not. So suppose that 'b' also denotes an imaginary man: f and b are still distinct. Why? The situation is essentially the same as that concerning Holmes and Pegasus, which we discussed in 4.5. Consider a closed world that realizes the representation concerning f. In any such world, f is fat and in the doorway, $Ff \wedge Df$; in particular, Df. Does the representation entail that both f and b are in the doorway, $Df \wedge Db$? Clearly not.[12] Nor is there any fact about existent objects that determines otherwise. So by the Principle of Freedom, there will be worlds of this kind where $Df \wedge Db$ fails, and so where Db fails. By the criterion of identity, f and b are distinct.

Context, then, makes all the difference. As for Quine's other questions:

- How do we decide?

[12] Verify this with a formal counter-model if you doubt.

We determine whether non-existent objects are the same or different in exactly the same way that we determine whether existent objects are: they are the same if they have the same properties in all (closed) worlds.

- How many possible men are there in that doorway?[13]

None. Being in the doorway is an existence-entailing property. So no non-existent object can have it.

- Are there more possible thin ones than fat ones?

No. There are zero of each.

- How many of them are alike?

All of them are alike (since there aren't any!).

- Or would their being alike make them one?

No. To be alike is to share some properties, maybe the important ones. But this is not sufficient for sharing all properties, and so being identical.

- Are no *two* possible things alike?

Of course two possible things can be alike, that is, share many important properties. Consider Tweedledum and Tweedledee.

- Is this the same as saying that it is impossible for two things to be alike?

No. To say that something, x, is possible is to say, roughly, that its properties do not violate the laws of logic. Let us express the thought that x is possible by $\blacklozenge x$.[14] Then the statement that no two possible things are alike is of the form: $\neg\mathfrak{S}x\mathfrak{S}y(\blacklozenge x \wedge \blacklozenge y \wedge x \neq y \wedge Axy)$. The statement that it is impossible for two things to be alike could be either of the form $\mathfrak{A}x\mathfrak{A}y(x \neq y \supset \neg\Diamond Axy)$ or $\neg\Diamond\mathfrak{S}x\mathfrak{S}y(x \neq y \wedge Axy)$. Both of these are quite clearly distinct.

- Or, finally, is the concept of identity simply inapplicable to unactualized possibles?

[13] The question is, of course, different from 'how many men is it possible for there to be in the doorway?' To which the answer is about three or four, depending on how big the men are.

[14] It is tempting to define $\blacklozenge x$ as $\Diamond Ex$. But this will not do. Abstract objects, like the natural numbers, are possible objects; but it will turn out in Ch. 7 that they are necessarily non-existent.

No. The concept is perfectly applicable. Two non-existent objects are the same if they have the same properties in all worlds (in the sense explained).

- But what sense can be found in talking of entities which cannot meaningfully be said to be identical with themselves and distinct from one another?

Statements of identity between objects are always meaningful. (But depending on what one thinks about vague objects, they may not always be true or false.)

We see, then, that Quine's rhetorical challenges pose no problems.

5.6 Conclusion

It is largely Russell and Quine who have brought Meinong, and more generally noneism, into disrepute. If anything, it is Russell's own view of the *Principles of Mathematics* that should have been brought into disrepute. Whilst the correct account of characterization was unclear, characterization was always going to provide a difficulty for Meinong and noneism. But the last chapter has now resolved that issue. And as for the other arguments that Russell and Quine deploy, so often thought to be fatal for noneism—as we have seen, these are groundless.

At a few places in this chapter, questions concerning abstract objects, platonism, and related issues have surfaced. We will turn to these issues in Ch. 7, together with some less standard objections that might be thought to tell against noneism. Before we do that, however, it will be useful for contrast to consider everyone's favourite example of (putatively) non-existent objects: fictional objects.

6
Fiction

6.1 Introduction: Fictional Objects

Non-existent objects, like existent objects, are of different kinds. This chapter is about one of them. Noneism, as such, does not commit one to any particular view about which objects do not exist. But the paradigm examples, about which nearly all noneists would agree, are fictional objects. Some of these, such as Holmes and Zeus, have already featured in the discussions of previous chapters. In this chapter, we will take a closer look at them.

In the first part of the chapter, I will look at their properties. In the second, I will look at a number of objections to the noneist account of fictional objects presented.

6.2 Fictional Operators

Fictional objects are those that feature in stories, plays, operas—and, we may add, myths and legends. Fictional objects, in this sense, may well exist. Thus, Napoleon features in Tolstoy's *War and Peace*, and Napoleon was a very existent character. Another story, *Sylvan's Box*, which illustrates the point is to be found in an appendix to this chapter. This is a short story about Richard Sylvan, who certainly existed as well. But most fictional objects do not exist. We may call fictional objects that do not exist *purely fictional*.

Fictional objects, whether pure or not, can clearly be the object of intentional states. We can think about them, feel sorry for them, think them funny, and so on. And the existential status of a fictional object is quite immaterial in this regard. We may or may not know this status; we may even be wrong about it. Such is completely immaterial phenomenologically.[1]

What of the non-intentional properties of purely fictional objects, however? It is tempting to suppose that fictional objects have the properties they are characterized as having. Thus, Holmes lives in Baker St, Zeus on Mt Olympus, and so on (and in *Meinong's Jungle*, Routley does seem to suggest this view sometimes). But the view should be resisted. For a start, such objects cannot actually have all the properties they are characterized as having. In the Doyle novels, for example, Holmes is certainly an existent detective. But he does not exist. Even if we were to distinguish between genuinely characterizing properties and others, as most meinongians do, the line still cannot be maintained. For a start, many of the characterizing properties of purely fictional objects are existent-entailing. Thus, in *Sylvan's Box*, Nick—who is not purely fictional—at one time, holds a certain box—which is purely fictional. You can't hold a box unless both you and the box exist. So it cannot be literally true that Nick held the box. Finally, and in any case, to suppose that such statements are literally true is to engender contradictions. Thus, in reality, Nick held no such box. So if it were literally true that Nick held the box, then it would be literally true that he both held and did not hold such a box. Now, whatever arguments there are for supposing that some contradictions may be true, there is nothing about noneism, as such, that requires this. (For all that is said in this book—at least to this point—the actual world may be quite consistent.) It would be wrong to saddle noneism with this extra feature.

In fact, the correct understanding of the matter is already explicit in Chapter 4. The objects of fiction, art, myth, and so on, are characterized in certain ways. And they have the properties they are characterized as having—and the consequences of these—not necessarily at the actual world, but at the worlds that realize the way the story, myth, etc., represents things to be. Thus, in reality, Nick held no such box, but, in the way that I represent things as being in *Sylvan's Box*, he did.[2] Thus,

[1] Walton (1978) argues that one cannot really admire (or fear, etc.) an object if one believes it not to exist; it is just a case of make-believe. But even for him, there is no problem about admiring or fearing a non-existent object if one does not so believe. Presumably, also, for him, you cannot admire an existent object if you believe it not to exist. Thus, existential status is irrelevant.

[2] In practice, and in the vernacular, the intentional operator is often suppressed, and understood contextually. Thus we say simply 'Holmes lived in Baker St' rather than the

if Φ is the appropriate intentional operator (see 4.2), we have something of the form $\neg A \wedge a\Phi A$. And it is easy enough to construct an interpretation where, for example, $@ \Vdash^{+} \neg Eb \wedge \neg Pb \wedge a\Phi Pb$, even where P is an existence-entailing predicate.

It is not, however, a trivial matter to say what the representations relevant to a fictional object are. The fact that something is not explicitly mentioned in a work of fiction does not mean that it is not part of the representation. Thus, an author will take for granted in the fiction certain things that carry over from actuality. As is clear when you read it, for example, *Sylvan's Box* takes place in Australia. This is not stated, however. What is actually stated is that it takes place in Bungendore, near Canberra. Neither is it stated that Bungendore and Canberra are in Australia. This is simply assumed. And since representations are logically closed, we are left to infer that the events take place in Australia.

It might be thought that an author's explicitly saying that something is sufficient for its being part of the representation; but even though this is normally the case, it is not universally so. A sophisticated author can state things that eventually turn out to be lies (in the fiction). Thus, a narrator may be telling a story about him- or herself, and it is left to the reader to figure out that the narrator is lying sometimes.

Giving a decent account of what, exactly, is (part of) the representation provided by a story or other work of art, is a difficult project. Fortunately, it is not one that needs to be addressed here.

6.3 Creating Objects

It is common to talk of authors of fictions as creating fictional objects. Thus, it might be claimed, Doyle created Holmes when he wrote the stories in question. It is natural to restrict this view to purely fictional objects. Thus, for example, I most certainly did not create Sylvan when I wrote *Sylvan's Box*. But even here, it might be suggested, I created a purely fictional Sylvan (distinct from the actual Sylvan).

Understood in some ways, the talk of creation is quite unproblematic. Doyle literally created the manuscripts of his stories; these manuscripts contain novel descriptions of people, crimes, places, even novel names like

correct 'In the way that Doyle represented things in the Holmes stories, Holmes lived in Baker St'.

'Sherlock Holmes'. (Here, I am talking about word tokens, not types.) If this is what it means to create a purely fictional character, so be it. But did Doyle literally create Holmes? More generally, are non-existent objects the creation of the cognitive agents who imagine them, fear them, worship them, and so on?

There is a problem about even how to ask the question. It would be natural to understand the claim that Doyle created Holmes as saying that Doyle brought Holmes into existence. (In the same way that—in the fiction—Frankenstein literally brought the monster into existence.) But this is not right. Holmes does not exist, and so Doyle did not bring him into existence. We might try to put the point counter-factually. Holmes would not have existed had Doyle not written his stories. This raises the question of how conditionals of this kind are to be understood. Roughly speaking, such a conditional is (actually) true if in those worlds that are the same as ours except that things are modified in the simplest way that realizes the antecedent, the consequent holds. The technical details of such an analysis are difficult and contentious.[3] But I will not go into them here; an intuitive understanding of these conditionals will suffice. So is it the case that in those worlds which are much like ours except that Doyle did not write his stories Holmes does not exist? Yes; in worlds where Doyle did not write, say because he died at birth, Baker St and its inhabitants would have been much as they were in this world. In particular, there would have been no Sherlock Holmes there.

But though the counter-factual is true, it fails to capture the intended connection between Holmes's status and Doyle's activities. For the status of Holmes in those worlds is exactly the same as in this: he does not exist. Any way of asking the question of creation has, therefore, to do so without deploying the notion of existence. The obvious candidate is: if Doyle had not written his stories, would something have been Sherlock Holmes ($\mathfrak{S}x\, x = h$)? The answer to this is 'yes': in the worlds where Doyle died at birth, something is Sherlock Holmes—Sherlock Holmes. Holmes is self-identical in those worlds, just as much as in this.[4]

[3] Impossible worlds are certainly required to handle conditionals with logically impossible antecedents. For a discussion see Priest (forthcoming *b*). Formal details are essentially those of Priest (2001), 10.7.

[4] Actually, this claim can be resisted if we take it that different worlds have different domains, and that Holmes is not in the domain in question. Thus, one might suppose that the domain of a world is dependent on the activities of some of its denizens—maybe the ones that exist at that world. But I am not inclined to this view. Consider some book that no one will ever

So if Doyle's activities did not determine Holmes's status, what was it that Doyle did to Holmes? Simply, Doyle was the first to imagine Holmes, and indeed, to give the character imagined that name, which we now use to refer to him. That is, he was the first to bear that particular intentional relation to him, in virtue of which we now imagine Holmes. (We are not, now, a million miles away from the question of how names for non-existent objects refer. I will take that issue up in the next chapter.)

What does it take to imagine an object? When Doyle formed the general intention to write about a detective of a certain kind, he had not imagined him; by the time he had written the first story, given him a name, and so on, he had. Imagining, like most humans achievements, may be vague. Suppose that I fear the man next door. When did that fear start: the first time I heard of him?—the first time someone told me something nasty about him? With most fears, there is no datable time at which they start. The situation can move from one of recognizing the object in question, through, maybe, having vaguely uneasy feelings, as one learns more, to, in due course, breaking out in a cold sweat. As with all such vague notions, there is a certain amount of arbitrariness in where one says the fear—or, in our case, the imagination—starts.

Once Doyle had imagined Holmes, however, he could go on and represent him as doing more and more things in different stories. It might be suggested that, in doing so, by saying more and more about Holmes, he was changing the characterization, and therefore speaking of a different object. But this would be wrong. Imagining new things about an object does not change the object in question. When I imagined Sylvan as possessing a box in *Sylvan's Box*, it was still Richard that I was imagining. When Holmes was first imagined, Doyle represented him to himself in a certain way. Holmes had his characterizing properties in those worlds that realize the way Doyle represented things as being in the story. But that representation was incomplete in all sorts of ways. When the first story (which was not the *Hound of the Baskervilles*) was written, there were some worlds that realized that story and in which Holmes became acquainted with the Hound, and some worlds of this kind in which he did not. As more and more stories were written, the class

write. This is about a pink axolotl called Zoe. It remains the case that someone could write this book, so Zoe could be in the range of quantification. But to say of something that it could be in the range of quantification is possible only if it already is.

of worlds gradually became further constrained by the representation (though never to a single world). But it was still the same Holmes that the stories were about.

The class of worlds can even bifurcate. Suppose that you and I decide to write a story about Holmes (Doyle's character). But our stories, whilst presupposing all that Doyle said, are incompatible. In my story, Holmes has a maiden aunt; in yours, not. Then in the worlds that realize the way that I have represented Homes, he has a maiden aunt. In the worlds that realize the way that you have represented him, he does not. Different worlds, but still the same Holmes.

6.4 Some Objections

I now turn to four issues that may be thought of as objections against the foregoing noneist account of fictional objects.

Issue Number One. The account given above allows explicitly for inconsistent fictions. Thus, a novel may be explicitly inconsistent. It is then realized at inconsistent worlds. But inconsistent fictions cannot be coherent. Where inconsistencies arise, we must override the author's say-so. In particular, we must chunk the representation into something like maximally consistent parts, and think of each of these as a version of the story.[5]

Now, it is certainly true that we may sometimes override the author's say-so. And it may well be the case that this is the correct strategy for inconsistencies that have entered the text by accident or by oversight.[6] Notwithstanding this, it is quite possible to have a story that is inconsistent, and essentially so. *Sylvan's Box*, as related in the appendix to this chapter, is a story that is inconsistent. But the inconsistency is no accident; it is essential to the plot. In particular, anyone who misapplied the principle of charity to interpret the story in a consistent way would have entirely misunderstood it. And its essence is entirely lost in any (collection of) consistent parts of it. Yet it is a coherent story. There is a determinate plot: not everything happens in the story; and people act in intelligible ways, even when the inconsistent is involved.

[5] An argument of this kind can be found in Lewis (1978). See, esp. pp. 274 f., 277 f. (Page references are to the reprint.)

[6] In fact, the Holmes stories are like this. Watson had an old war-wound. In one story Doyle says that this was in his leg; in another, that it was in his arm.

In fact, we can extract more from this example. Representations, as I have noted, are closed under a logic. The logic in the case of *Sylvan's Box* is paraconsistent; that is, a logic in which contradictions do not imply everything. Certain inconsistencies hold in the story, but not everything does. I would like to take this as an argument to the effect that a paraconsistent logic is the (uniquely) correct logic: we certainly have no temptation to apply Explosion ($A, \neg A \vdash B$) and infer that everything happened in the story. But though I do think that Explosion is invalid, I do not think that the example establishes it to be so. Probably, for any logic, one could write a story for which the correct canons of inference were given by that logic. Thus, take, for example, quantum logic—in which distribution ($A \wedge (B \vee C) \vdash (A \wedge B) \vee (A \wedge C)$) fails. I think that it would be easy enough to fashion a story in which Plank's constant was, say, 143, and, accordingly, macroscopic objects moved in quantum fashion. Thus, I might arrive in a room having gone through one of two doors without having arrived in the room having gone through one or other of them. It seems to me that, though the correct (paraconsistent) logic is the default logic for reasoning about a fictional situation, one can, in some way, moderate this in such a way as to accommodate the demands of the particular story. How one does this, though, is another matter.

Issue Number Two. It is frequently noted that fictional objects are incomplete in a certain way. Thus, it is neither true that Holmes had a sister nor that he did not. (Doyle never tells us.) If one supposes that 'fictional truths' are literally true, this kind of incompleteness can be handled by supposing the actual world, @, to have truth value gaps. But there is a more virulent kind of incompleteness (noted, in effect, in 4.4). For Holmes, as are all people, was either left-handed or right-handed (or ambidextrous). But he was neither right-handed nor left-handed. (Doyle never tells us.) Even if some sentences are neither true nor false, we cannot have a truth of the form $A \vee B$ when neither A nor B is true. Hence the semantics is incorrect.

But this problem is solved on the above account. As Doyle represented things as being, Holmes was either left- or right-handed, but he did not represent him as left-handed and he did not represent him as right-handed. Thus, it is easy enough to construct an interpretation where $@ \Vdash^+ d\Phi(Lh \vee Rh)$, but where neither $@ \Vdash^+ d\Phi Lh$ nor $@ \Vdash^+ d\Phi Rh$. (The denotation of h is in the extension of either L or R at every world, w, such that $@R_\Psi^{\delta(d)}w$; but in some it is in the extension of L but not R, and vice versa.)

The same technique handles incompleteness of the more mundane kinds too. Interestingly, then, although we need inconsistent worlds to handle inconsistent fictional objects, though we have incomplete worlds (i.e. worlds with truth value gaps) in the semantics, we do not need them to accommodate 'incomplete' objects.

Issue Number Three. The account accommodates the truth of other things that we would naturally want to say about fictional objects. For example, let a be any actual detective. Then it is true that Holmes (h) is more famous than a. That is, more people have heard of Holmes than of a—where, here, 'hear of' is an intentional predicate. This sentence cannot be formalized in the language we have used, since it has no quantifier 'More xs are . . . than . . .', it would be easy enough to extend the language with such a quantifier and model this assertion; but even without this, one can construct a simple model where more people satisfy the formula xHh than xHa at $@$. (Here, H represents 'has heard of'.)

But what about truths concerning fictional objects whose predicates appear to be existence-entailing? For example, Tolkein tells us that Bilbo Baggins, being a Hobbit, was short. Priest is 6 ft. 4 in. It would therefore appear to be true that Priest is taller than Bilbo. But *being taller than* would certainly appear to be an existence-entailing predicate, so this would seem to be false. We may accommodate its truth in the following way. There are numbers, x and y, such that x is Priest's height; in the world as Tolkein represented it as being, Baggins's height is y; and $x > y$. It is easy to construct an interpretation in which $@ \Vdash^{+} \mathfrak{S}x\mathfrak{S}y(Px \wedge t\Phi By \wedge x > y)$.[7]

Similar examples can be treated in the same way, though there may be a touch more artifice involved. For example, we might want to say that a fictional character was more angry (in some fictional context) than some actual character (in some real context). It is less natural to talk of some literal degree of anger than of some degree of height. But, it seems to me, it is quite possible, none the less.

Issue Number Four. Here I take up the point I flagged in 4.3. We saw there that if the way that things are represented as being is realized by the actual world, and if there is a unique object there satisfying the characterizing

[7] Actually, things are not quite so simple, since Tolkein specified no particular height for Baggins. Rather, there is a range of heights such that, in the *Hobbit*, Baggins has a height in that range, and each height in the range is less than Priest's height. But this extra complexity changes nothing essential here.

conditions, then that is the object characterized. But now suppose that I tell a story. It is about a man called Napoleon, who does certain things. Suppose that it is a story with a historical setting, but, as far as I know, entirely fictitious. But suppose, also, that it turns out that the story I tell just so happens to be realized by the real world, and that there is a person who does the things that Napoleon does in my story—including, if you like, being called 'Napoleon'. It follows that the Napoleon of my story is that person. This might be thought odd. Any similarity between the Napoleon of my story and the actual character is entirely fortuitous. Maybe, though, one should just live with the conclusion. I certainly did not intend the story to be about him; but then things often happen that one does not intend.

There is another possibility here, though. When the nineteenth-century astronomers presented a theory about Vulcan, they certainly intended it to apply to the actual world. But when I tell a work of fiction, I deliberately intend to exclude this possibility. Thus, we should perhaps rule out this world as a candidate for satisfying it. It therefore makes sense to suppose that the appropriate intentional state involved in representing things in this case, Φ', is different from that, Φ, in which I intend the story to be veridical. We may take $a\Phi' A$ to be something like: 'a represents A as holding non-actually [in the matter at hand]'. For Φ' we do not have $@R^{\delta(a)}_{\Phi'}@$. Since the actual world is not one of the worlds realizing the situation represented, we are not forced to conclude that the name 'Napoleon' refers to the actual Napoleon.

6.5 Conclusion

In this chapter, we have looked at one important kind of non-existent object, purely fictional objects. We have seen how a noneist account of fictional objects works, and that it does not fall to a number of the more obvious objections. Fictional objects do not exhaust the menagerie of non-existent objects, however. Mathematical and other abstract objects can be taken to form another large class. In the next chapter, we will look at these. This will occasion the consideration of a number of further objections to noneism, some of which are relevant to fictional objects too. But first, the short story which illustrates a number of the points made in this chapter.[8]

[8] Some background to the story. Richard is Richard Sylvan (né Routley), who died suddenly in June 1996; Nick is Nick Griffin, his literary executor; and the visit to Bungendore actually

6.6 Appendix: *Sylvan's Box*

I still couldn't really believe it was possible. Richard dead. Never again would I see him. Never again would we talk, share ideas, problems, a bottle of wine. I changed down to overtake a slow-moving car in front of me. If he had been 80, the victim of creeping senility, reduced to a rocking chair, it would have been different. But his vigour had been palpable. He was still in full cry, working on his numerous philosophical projects. The ultimate Australian iconoclast. I changed back up, and moved over as I passed the car. Then there was his hobby. Building houses. I don't mean designing houses. I mean physically building them. Every day he would write from about dawn till lunch-time. Then he would go and carry bricks, move beams, dig foundations. It just didn't seem possible that this body had a heart that could give out. Suddenly. Just like that. But it had.

The journey from Canberra to his farm at Bungendore was a short one, at least compared to the 1,000 kilometres I had recently covered from Brisbane. The sun was going down, and it was the magic time of day. That time when the sun mercifully elects to hide itself for a few hours, and the roasted earth heaves a sigh of relief. The twilight hues softened the Australian bush that Richard loved so much. And the colours of the sunset—pinks, lilacs, peaches—were reflected in the calm waters of Lake George. I would have to hurry, or it would be difficult to find—let alone follow—the three kilometres of rough dirt track that led to the farmhouse.

By the time I found it, the light was almost gone. The headlights showed the boulder-strewn track that the poor little car would have to negotiate. The suspension didn't know what was about to hit it. The drive up the track seemed interminable. Often I thought that I had taken the wrong turn and ought to turn back; but eventually, on reaching the top of a particularly nasty stretch of track, I was rewarded with twinkling lights—driven by the massive solar-powered batteries under the house.

Nick had heard the car, and came out to meet me. 'Hi, Graham.' 'Hi, Nick.' We shook hands and looked at each other, sharing thoughts that neither of us needed to express.

took place in December of 1996. The story, needless to say, is not an accurate description of my visit to Bungendore, but it is something like the actual visit minimally modified by the finding of the box—except for the end, anyway. The Zenoesque arguments mentioned in the story, to the effect that changes may realize contradiction in the world, can be found in chs. 10 and 11 of Priest (1987). The discussion of quantum mechanics referred to by the fictional Nick is on pp. 376–9 of Priest, Routley, and Norman (1989).

'Come in. Have you eaten?'

'Not for some hours,' I said.

'Right, I'll get something. I haven't eaten myself yet. It'll be a bit primitive, I'm afraid. I haven't mastered the wood-fired stove yet. Put your bag in the bedroom. I'll sleep on the couch in the lounge.'

The place was almost entirely as Richard had left it, expecting to return: the books, the small collection of wine that he enjoyed, the tools. All that was missing was Richard; and it was hard to believe that he was not about to walk in, after having been working on the property, and in his gruff but gentle voice say, 'So you found it all right? Want a glass of wine?' But he wasn't.

There was one thing that was different about the place, though. Nick had been busy. He had been working on Richard's papers, many of which were now stacked up in boxes, or simply in piles. In fact, the only way to get around was literally to traverse a path through Richard's intellectual legacy.

'Please excuse the mess,' said Nick in his very English way, 'but there is just so much of the stuff. The only way I have been able to get it into any order is to make a pile of the material for each of the topics that Richard was working on—after first noting where I found each thing. Richard's filing system was . . . shall we say, original.'

Despite his protestations, Nick prepared an excellent meal. We also opened a bottle of Richard's wine and drank to his memory.

'Thanks for coming down,' said Nick.

'I'm happy to do what I can,' I said, 'You are the one that is doing all the hard work. It's a hell of a job.'

'Well there's certainly a lot more here than I had thought,' he said, 'and I've so little time before I must go back to Canada. I haven't really had time to look at any of it. I'm just trying to get things documented, and in a fit state for people to work on. That pile over there is all on environmental ethics, there seem to be at least two books on the go there; that pile contains the working notes for a book on pluralism; that pile is all on paraconsistent logic, I don't really know what's there; the pile next to it contains stuff on Meinong and associated topics. Then that big mess over there is all on relevant logic. There are parts of Volume II of *Relevant Logics and their Rivals*, which Richard had been sitting on for years, and part of a newer book on the applications of relevant logic. The other piles are more of a mixture: correspondence, travel documents, building plans. And I've only just made a start on the stuff in his office at the ANU.'

'Right,' I said, and surveyed it all. 'And what do you want me to do while I'm here?'

'Well, I'm no logician, so I'd like you to look at the logic material and tell me what you think is there, what might eventually find its way into print, who might work on it. That sort of thing.'

The thought of doing this was somewhat overwhelming at this time of night. And in any case, the day's driving and the wine were taking their toll on me. I told Nick that I would start in the morning, and wandered off to bed. Nick settled down for the night shift. I was asleep within minutes.

I was awoken by the sun. Richard hadn't believed in curtains. I thought that I might as well get on with things. Nick was still sound asleep on the sofa, so, as quietly as I could, I made myself a pot of tea, and moved a couple of boxes onto the verandah. It was a beautiful day, warm but not yet hot. The sun shone on Lake George, which lay below us. The smell of the bush, mixed with the taste of the tea, both calmed and invigorated the senses at the same time. I started to go through the boxes. These had come from the pile Nick had described as 'relevant logic'. It was a real mixture: draft chapters, notes, papers by other people with annotations, the occasional table of contents. Examining all the material was rather slow going. Someone was going to have to do an *awful* lot of work on it.

A couple of hours or so later, I heard Nick moving around in the house, and went back in. We had some breakfast and discussed what I had learned from the material.

'It's going to take most of the day to get through the material,' I said.

'Okay, well I'm going to drive in to Canberra and get on with some things there,' he said, 'Will you be okay here?'

'Sure.'

By the time that Nick left, it was already too hot to work outside any more. I decided to leave the rest of the material on relevant logic for later, and have a look at the material on paraconsistency. Being smaller in quantity, it might be more manageable. There was the same mixture of notes and papers as in the other pile, but here there was also correspondence from me. I had forgotten all that. Memories came flooding back. My mind wandered off, reliving the past; all the times we had spent together; all the discussions we had had; all the joint work. Before I realized it, it was lunch-time. I made myself a sandwich, and carried on with the papers. It seemed to me that there wasn't anything very new in this particular pile. Just the record of past work.

As I was putting the last batch of papers back, I noticed a small box located between that pile and the one on Meinong. It was too small to have papers in, I thought; maybe it contained some more letters. I picked it up and examined it. It was of brown cardboard of poor quality, made in a developing country, perhaps. The lid was taped down, and on it there was a label. In Richard's own handwriting—under which dozens of typists had suffered over the years—was written 'Impossible Object'. 'Well, that explains the ambiguous place in Nick's categorization,' I thought. There was very little else on the box except some print on one side. It was very faded, and even more difficult to make out than Richard's handwriting. Just barely, I thought, I could perceive a date. Maybe 1979.

Carefully, I broke the tape and removed the lid. The sunlight streamed through the window into the box, illuminating its contents, or lack of them. For some moments I could do nothing but gaze, mouth agape. At first, I thought that it must be a trick of the light, but more careful inspection certified that it was no illusion. The box was absolutely empty, but also had something in it. Fixed to its base was a small figurine, carved of wood, Chinese influence, south-east Asian maybe.

I put the lid back on the box, and sat down hard on an armchair, my mental states in some disarray. I focused on the room. It appeared normal. My senses seemed to be functioning properly. I focused on myself. I appeared normal. No signs of incipient insanity. Maybe, I thought, it was some Asian conjuring trick. Gently, I reopened the box and gazed inside. One cannot explain to a congenitally blind person what the colour red looks like. Similarly, it is impossible to explain what the perception of a contradiction, naked and brazen, is like. Sometimes, when one travels on a train, one arrives at a station at the same time as another train. If the other train moves first, it is possible to experience a strange sensation. One's kinaesthetic senses say that one is stationary; but gazing out of the window says that one is moving. Phenomenologically, one experiences what stationary motion is like. Looking in the box was something like that: the experience was one of occupied emptiness. But unlike the train, this was no illusion. The box was really empty and occupied at the same time. The sense of touch confirmed this.

Again, I put the lid on the box; I put it down. Then I wandered off to see if Richard had a bottle of Scotch. It seemed that he didn't, so a large mug of tea was the best I could do. Probably just as well. I sat sipping the tea for some time, rapt in thought. What I had discovered seemed so unlikely, impossible even—just as the box said. But there are many things in Heaven and Earth that are not dreamt of in your philosophy—Horatio. No doubt the thought that the earth, the most stable and solid thing in our experience (except for the odd quake and tremor), is spinning through space, must have been equally hard for people in the sixteenth century to get their heads around. Goodness knows what Newton would have made of time running at different rates—maybe even backwards—with respect to our frame of reference. And *I* should be the last person in the world to be shocked by this particular discovery.

My thoughts were interrupted by the sound of Nick arriving in the Land Rover, and I realized that I was sipping cold tea. For some reason I panicked. What on earth was I going to say to Nick? All too quickly, he was coming through the door, another large batch of papers under one arm and some provisions under the other. 'Hi,' he said.

'Hi Nick,' I said, trying to appear as normal as possible.

'Had a good day?' he said. He looked at me quizzically: 'Are you all right? You look a bit pale.'

There didn't seem to be any point in dissembling. 'Nick—have you found anything, er . . . odd, amongst Richard's things?'

'Well, can't say that I have. A few spiders used as bookmarks. That sort of thing. Is that what you mean?'

'Not exactly.'

I went over and pointed at the box, now nestling comfortably between paraconsistency and Meinong again. 'Do you know what's in this?' I said.

'No. I wasn't sure what it contained, so I just put it in an appropriate place until I had time to get back to it.'

'Sit down,' I said, and handed him the box. He looked at it, looked at me, and then started to remove the lid. My heart beat wildly. Maybe I was just about to appear an enormous fool; maybe he was just about to have me taken away and certified.

I watched him closely. In the space of a few seconds, his look turned from curiosity, to incomprehension, to sheer disbelief, mixed, I thought, with a little panic. That must be exactly what I had looked like. For some moments he stared at me, unable to say anything. After what seemed a rather long time, I managed to say rather feebly 'Odd, isn't it'.

'Yes, I've never seen anything quite like it', he replied, with the sangfroid that had won the English an empire.

'Not bad', I thought, 'for someone whose *lebenswelt* lies in tatters.'

I left him alone for a while, so that he could put enough of it back together to at least have a sensible conversation. Maybe a walk in the bush, in what was left of the day's light, would help to put its events into some perspective.

When I got back, Nick was putting the finishing touches to a meal. 'Let's eat', he said. We ate. There were a lot of silences at first, but soon the discussion started to flow. And then there seemed no stopping it. It flowed through many hours and bottles of Richard's wine. We talked about where the box had come from; we talked about why Richard had never said anything about it to either of us—or anybody else as far as we knew; we talked about what the existence of such an object meant, for logic, for metaphysics; we talked about how one might construct such an object. The discussion was, it must be said, on all counts completely inconclusive. Except one: we figured that Richard had probably acquired the box around, or just after, the date on its side, and that it was likely that he had picked it up on one of his many trips overseas. Richard had been a great traveller. There weren't many places that had universities—or that might even *conceivably* have had universities—that Richard hadn't visited. We guessed that he probably found it in Indonesia or Malaysia.

By the small hours of the morning we were talked out. Nick said that the day had had quite enough experiences in it, and headed for the couch, finishing the last of a fine Hunter Valley Shiraz that neither of us was now in a position to appreciate properly. I didn't feel like sleeping, so I took the pile of Richard's papers that Nick had designated 'on Meinong' into the bedroom, and began to work through them.

I was hoping to find something that would help answer the questions over which we had spent so long getting nowhere. By and large, I was disappointed. The papers contained chapters of *Meinong's Jungle*, notes for some essays on existence and impossibility—very old ones—some more correspondence with various people, including me; but nothing that cast any light on the issues of the moment.

The old letters did jog one relevant memory, however. When I first met Richard, we had disagreed over whether the actual world could contain contradictions. I thought that maybe it could. He thought that it was only non-existent things (such as propositions and mathematical objects) that could be inconsistent, that contradictions were all 'off-T' as he was fond of putting it ('T' being his name for the actual world). We had argued about this on and off for some years. He had never been able to persuade me that there was any reason why existence should imply consistency. I, on the other hand, had never been able to convince him with my arguments—largely stolen from Zeno—that things in a state of change, actual things, realized contradictions. But his attitude had changed in the early 1980s, very suddenly and for no reason that I could fathom. When I asked him why, all he had said was 'Maybe you are right.' Naively, I had put this down to the attrition of my arguments. Now it seemed to me that there was a much more likely explanation. He had held the proof in his hands.

But this made another question even more puzzling. Why had Richard never said anything to me about it? Richard never minded admitting that he had been wrong—on the rare occasions that one could show it. And the box gave enough material to confound the opponents of paraconsistency once and for all. So why the silence? As I pondered the issue again, the first rays of the morning light started to appear over Lake George and filter into the bedroom. Outside, the tenebrous shapes of the trees acquired a tinge of iridescence. I drifted involuntarily into sleep.

When I awoke, the sun was high in the sky. I took a cold shower and collected my thoughts. I had already stayed longer than I had intended. I needed to be back in Brisbane by tomorrow, and would have to leave today. I collected the few things that I had brought with me, together with some notes by Richard on impossible worlds that I thought I might be able to edit into a finished article, and loaded them into the car.

When I returned to the house, I found Nick working on a pile of papers tucked in the corner of the room. 'I think I know how it might work,' he said.

'What?' I replied.

'You know, the box.'

'Really? Tell me.'

'Well it's only a rough idea, but it may be on the right track. I recalled something that you and Richard wrote many years ago. I think it was in one of the essays in *Paraconsistent Logic*. It was about quantum mechanics—and particularly, the two-slit experiment. Given the set-up, a particle seems to do the impossible: go through both of two distinct slits simultaneously. You suggested that this is exactly what it *does* do. At the micro-level inconsistencies can actually be realized. Well, suppose that that's right. And suppose that some way could be found to bring about the same effect at the macro-level. I don't really know how. Maybe it's a bit like Schrödinger's cat. A macro-state is made to hold in virtue of some quantum event. But rather than the cat being dead and alive, the box is both occupied and empty.'

'That's it?' I said, 'It's a bit thin.'

'Oh', said Nick, a little deflated, 'Got any better suggestions?'

I had to admit that I hadn't. Nick's idea was a bit wacky, but any explanation, I reflected, would have to seem like that.

I reminded Nick that I would have to leave soon. As we ate breakfast the imminence of my departure raised a new question, one that, rather amazingly, neither of us had thought to ask till now: 'What are we going to do?'

We immediately agreed that we should of course make the box public. It would be of major importance for logic, metaphysics, and, if Nick was right, physics too. And that was putting it mildly. There seemed to be no question about the matter.

But then we fell silent as each of us started to think about the possible consequences of such an action. As I spread marmalade on another piece of toast, I conjured up the images of life afterwards. The incredulity with which the announcement would be greeted. The probability of being branded as cranks by large numbers of the profession. The media attention it would certainly draw. All the prying of psychologists, journalists, real cranks. Life would be altered irreversibly. Both our professional and our private lives would be changed, one way or another; and not necessarily for the better. Many people crave fame and fortune, but those who obtain it often live to regret it. I suddenly understood why Richard had said nothing about the box. It would have destroyed the peace that he loved so much. The tranquillity of his farm in the bush, the solitude, the sun rising over Lake George, the singing of the birds.

And if Nick was right about the physics, this would just be the start of things. Who would take possession of the box? Whatever knowledge it yielded was the sort of thing that corporations would want to use to make enormous profits, that individuals would kill for, that governments would want to keep secret, that the military would want to use to make weapons with. The apple of knowledge has often acquired a sour taste for humankind.

But then, it was also possible that whatever physical mechanism underlay the box would be discovered in time anyway, at least if paraconsistent logic were ever taken seriously by the scientific community. What was the point of trying to suppress it now?

Neither Nick nor I had spoken for some minutes. I did not need to talk to him to see that the same thoughts had been going through his mind. A look into his eyes told the tale. Doubtless, a look into mine told the same tale. The box was on the table between us. We both stared at it forlornly, as if we hoped that it, itself, would give us an answer. In a way, it did.

We stood up. I carried the box outside; Nick carried the box outside. I opened the car door; Nick picked up a spade and dug a hole. I put the box in the car; Nick put the box in the hole. I closed the door on the box, and locked it; Nick covered the box with dirt and stamped it down. We turned to face each other. Silently, we embraced. I got into the car and drove off into a world that would never be the same for either of us again.

7

Mathematical Objects and Worlds

7.1 Introduction: Kinds of Non-Existent Object

Purely fictional objects are not the only kind of non-existent object. Arguably, another such class comprises abstract objects, and particularly mathematical objects. Not all noneists have taken these objects to be non-existent. As we noted in Ch. 5, for Meinong himself, and Russell in the *Principles of Mathematics*, these were not non-existent objects, but subsistent ones. However, Routley took abstract objects as non-existent.[1] And the picture of reality whereby it comprises the existent, which are concrete objects in space and time, and, for the rest, the non-existent, has an appealing cleanness about it.[2] In this chapter we will start by looking at the treatment of abstract objects as non-existent.

In previous chapters, much use has been made of the notion of worlds, their properties and relations; indeed, these have been central to the analysis of intentionality offered in the book; and a natural question concerns the status of worlds themselves. Worlds have a vexed status,

[1] Routley (2003). The first part of this paper is to the effect that mathematics is not only noneist, but non-extensional. I shall have little to say about this matter here. I note, however, that if a noneist account of mathematical objects is correct, then the treatment of the CP given in Ch. 4 reinforces Routley's position on the matter. For if mathematics is about non-existent objects, and if the behaviour of these cannot be understood without talking about worlds other than the actual, then other-worldliness is built into mathematics. But at least as understood in standard modern logical semantics, it is precisely the essential employment of such worlds that is the defining moment of intensionality (with an 's').

[2] As I indicated in the Preface to the book, this was not quite Routley's picture. For him, only concrete objects *presently* existing exist. Thus, Socrates and the end of the Earth do not exist. I will not follow him down this path.

even without noneism. But if one is a noneist, an obvious possibility is that all worlds, with the exception of the actual world, are non-existent objects. This was, in fact, Routley's view. Worlds as non-existent objects will be the subject of the next part of the chapter.

Purely fictional objects, abstract objects, and worlds are three kinds of non-existent objects. The categorization is not meant to be exhaustive[3]—or exclusive, for that matter—as we will have occasion to note. But they are some of the most important kinds; and our discussion of them will raise a number of questions for, and possible objections to, noneism; for example: How does one know about such things? Is noneism really platonism in disguise? In the rest of the chapter, we will look at these issues.

7.2 Abstract Objects

Abstract objects have a notoriously troubled locus in philosophy. Properties, relations, propositions, and, above all, mathematical objects, have an ontological and epistemological status that is highly problematic. All accounts seem to face difficulties.[4] Platonism of some form is, perhaps, the default position. And a noneist can certainly endorse a platonist account. Meinong himself subscribed to a form of this. For him, as we have already seen, abstract objects do not exist, but they do subsist; that is, they have a distinctive form of being. But for a noneist, a simpler view beckons; it certainly summoned Routley. Abstract objects are just another kind of non-existent object. This at least accounts for the fact that there seems to be a very great difference in kind between ordinary concrete objects and abstract objects. The difference between existence and non-existence is about as great as can be! The question is whether a noneist account stands up to closer inspection.

Let us start with the question of what, exactly, an abstract object is. The answer to this is by no means obvious. A natural first suggestion is that abstract objects are ones that do not enter into causal chains with you, me, and the things that we, in turn, interact causally with. Such an account is problematic, however. For example, a modal realist, such

[3] e.g. ideal objects in science, such as frictionless planes and perfect gases, form another plausible class.

[4] For a discussion of some of the problems, see Priest (1987), 10.4.

as Lewis, takes worlds other than the actual to satisfy this criterion. Yet these worlds are not abstract objects, but physical ones, just like the actual world. The noneist has a different problem. Purely fictional objects, like Holmes and Zeus, do not enter into causal chains with respect to us. 'x doing such and such caused y to do such and such' would appear to be an existence-entailing property. Since these objects do not exist, they cannot enter into causal relations; but they are not abstract objects, at least as usually conceived.

For a noneist, a major difference between standard purely fictional objects and abstract objects would seem to be in the *mode* of their existential status. Holmes and Zeus do not exist, but they could have done. There are possible worlds that realize the Holmes stories, and in those Holmes does exist. The status of being an abstract object, by contrast, would appear to be a non-contingent matter. To exist is to be concrete. Are there worlds in which, say, 3 is a concrete object? Yes, as we saw in Ch. 4, there are worlds in which anything can be realized. But the world in which one can hold 3 in one's hand hardly seems to be a possible one. Thus, we may take it that an abstract object not only does not exist, but necessarily does not exist.

This suggests taking the necessity of existential status to be the defining criterion of being an abstract object. This account, however, is also problematic, given what has already been said about purely fictional objects. The box in *Sylvan's Box* has contradictory properties. It is natural to suppose, therefore, that there is no possible world where it exists. Yet it is not an abstract object: it's a box (in the story). One may avoid this particular objection, as I would be inclined to, by simply accepting the fact that there are possible worlds at which contradictions are true. But the point is more general than this. Whatever one takes the correct logic to be, one can construct a story in which there are objects that have logically impossible properties. In 6.4 I noted, for example, the possibility of a purely fictional object, a person, who violates the logical law of distribution. The person satisfies the condition $A(x) \wedge (B(x) \vee C(x))$, but not the condition $(A(x) \wedge B(x)) \vee (A(x) \wedge C(x))$. This is a person (in the story), not an abstract object, but one who exists only at impossible worlds.

Perhaps more success can be had with a counter-factual criterion which, in effect, combines the previous two proposals: an abstract object is one such that, *if it did exist it would still not causally interact with us*. Conversely, a concrete object is one such that, *if it did exist, it would* *causally interact with us*.[5] Holmes, were he to have existed, would have entered into causal chains with us. We could have seen him entering and leaving his rooms in Baker St. He is, therefore, a concrete object. Sylvan's box, also, had it existed, would have entered into causal chains with us. Thus, for example, in the story, Nick holds it in his hands. Therefore, it also is not an abstract object. But consider an object that we would naturally take to be abstract, say the number 3. The simplest way to accommodate the claim that this exists is to suppose that the world is such as the traditional platonist takes it to be. If the noneist is right about abstract objects (so that they do not exist) and the existential status of abstract objects is necessary (so that they cannot exist) then such a world is not a possible world. Never mind; worlds of this kind realize the antecedent of the conditional, but it remains the case that 3 does not enter into causal relationships with us at such worlds. Platonists do not normally think that one can see or hold 3. Of course, there will be other, impossible, worlds where we do enter into causal relations with 3. There are worlds, for example, in which 3 is a cat, and so can be stroked. But these are much more bizarre than the usual platonic picture, and so not the worlds relevant to the counter-factual. Thus, if 3 were to exist, it would still not interact with us causally. That is, it is an abstract object.

The account, then, gives us a plausible understanding of what an abstract object is. It should be noted, though, that on this account an abstract object can be purely fictional. Thus, for example, suppose I tell a story about some (actually non-existent) object which is incapable of entering into causal relationships with us. This is a purely fictional object, but also an abstract object. If it were to exist, we would not be able to see it or hold it. Thus, the categories of abstract and purely fictional objects are not disjoint. But this is no problem: one can, after all, tell a story about 3, or about any other non-existent object, just as much as one can tell a story about an existent object, such as Sylvan. Note, also, that, on this account, the division between abstract and concrete objects is not exhaustive. There is no reason to suppose that one or the other of the conditionals: *if x did exist, x would causally interact with us* and *if x did exist, x would not causally interact with us* is true. What would happen if *x* existed might be indeterminate in this regard.

[5] For counter-factuals, see 6.3.

7.3 Worlds

Let us now turn to the status of worlds. One must distinguish immediately between the worlds themselves and the mathematical representation of them. In Part I of the book I gave a semantics for a language with intentional operators and predicates. The semantics is of a kind familiar in contemporary logic, and deploys the apparatus of set-theory. As such, the objects with which it concerns itself are mathematical ones; and the question of their status has been addressed in the previous section.

But worlds themselves are not the same thing: they are what the mathematical machinery represents. In a similar way, we may represent space and time (and the objects in them) by mathematical structures, such as the real line or Euclidean 3-Space. These are mathematical structures; space and time are not (at least, not in the same sense). If we apply the mathematical representation to space and time, we can do so because the two share a common structure. By figuring out the structure of one (the mathematical one), we can therefore learn something about the other. (I will have more to say about this later in the chapter.) In a similar way, the mathematical semantics of Part 1 of the book provides a representation, not of space, but of language, the extra-linguistic, and the relationship between them. And if it is correct, something must be represented by it, and a something that shares the relevant structure. The extra-linguistic wing of the relation includes the worlds themselves, their properties and relations.

What is the status of their being? As far as I can see, the preceding chapters are compatible with any answer one might wish to give to this question. One might, for example, be a realist of Lewis's kind,[6] and take worlds to be concrete objects of the same kind as ours, just not actual. Alternatively, one could take non-actual worlds to be abstract objects of some kind, such as sets of sentences, or constructions out of properties or universals.[7] Of course, if one subscribes to the account of abstract objects of the previous section, this will collapse into a noneist account; but it is quite possible for a noneist to give a different account of abstract objects, as I have noted.

One does not, of course, have to suppose that all worlds have the same status. The actual world is naturally thought of as a special case (though not necessarily, as modal realism reminds us). But one might suppose also,

[6] See e.g. Lewis (1986). [7] See e.g. Priest (2001), 2.5–8.

for example, that possible worlds and impossible worlds have different sorts of status. Thus, one might be a modal realist about possible worlds, but take impossible worlds to be abstract objects. However, I know of no good arguments for distinguishing between the status of different kinds of non-actual worlds in this way—just as there seems to be no good reason to distinguish between the status of physically possible worlds and that of physically impossible (but logically possible) worlds. A simple uniform policy therefore recommends itself.[8] And in the context of noneism, the obvious policy is to take all worlds other than the actual to be non-existent objects.

This does not necessarily mean that they are abstract objects. Indeed, if one applies the criterion of the last section then, at least for the most part, they are not. The worlds that realize the Holmes stories are replete with things that, were they to exist, would be standard physical objects, like people and hansom cabs. Were these worlds with their denizens to exist, we would be able to interact causally with them. Just as Lewis claimed, then, these worlds are just like the one in which we live—or they would be if they existed. There can, of course, be unusual worlds; for example, worlds where nothing exists. It is not clear that, if this world were realized, it would enter into causal relationships with us. Maybe, then, one should take such a world to be an abstract object. But at any rate, all worlds other than the actual have the uniform status of non-existence.

What of their properties? We have made use of a number of these in previous chapters. We have distinguished, for example, between possible, impossible, and open worlds. The relationship $\Vdash^{\pm}$—that is, the one represented by this relationship in the formal semantics, where the interpretation, $\mathfrak{I}$, in question is the veridical one—is a relationship between worlds and other things. To be specific, since statements concerning it are of the form $w \Vdash^{\pm}_{s} A$, $\Vdash^{\pm}$ is a three place relationship. The first argument is a world, the other two arguments, s and A, are a function and a sentence (type). These are abstract, indeed mathematical, objects. Now, none of the properties and relations at issue here is existence-entailing. To say that a world is possible, for example, does not entail that it exists, any more than to say that an object is possible does. Attributions of modal status are logical attributions, like statements of identity. And $\Vdash^{\pm}$ is not existence-entailing, either; what I have already said about abstract objects

[8] For further arguments against drawing an ontological distinction between possible and impossible worlds, see Yagisawa (1988).

delivers this fact for at least the last two argument places, and the first is no different. Thus, there is no problem about taking these properties of the worlds in question to be properties they have at the actual world.

7.4 Five Objections

Given the preceding discussion, we can now turn to some objections against noneism, and particularly against a noneist account of abstract objects. There are five natural ones that I will discuss, which are as follows.[9]

1. If some objects do not exist, they cannot enter into causal connections with us. How, then, can we refer to them or speak of them at all?—which a noneist obviously requires us to be able to do.[10]
2. Similarly, since they do not enter into causal connections with us, how can we know anything about them?—which we certainly can if noneism is right.
3. According to the account given, fictional objects and mathematical objects are of a kind: non-existent. Yet, they seem to be quite different sorts of things. For example, we can make up truths about fictional objects as we go along, not mathematical objects. So how can this be?
4. We often apply mathematics to tell us about concrete objects, like shopping, bridges, microchips. How can non-existent objects possibly tell us anything about things that do exist?
5. The noneist and the platonist hold that some objects do not enter into causal relationships with us. They disagree about whether or not they exist, though. But in the end this is just a difference of terminology. When the noneist says that something is an object, the platonist says that it exists; when the noneist says an object exists, the platonist says that it is concrete (and exists). The noneist is just, therefore, a platonist in disguise.

In the rest of this chapter, I will take up each of these points in turn.

[9] Some of these, and some other objections to a noneist account of mathematical objects, are taken up in Routley (2003).

[10] A version of this objection can be found in Walton (1990), 10.1. In the same section, Walton perpetuates the confusion to the effect that Meinong took non-existent objects to have being.

7.5 Referring

Non-existent objects do not enter into causal relationships with us, it is true: causation is, as we have noted, an existence-entailing relation. But reference does not require causation. This is obvious in the case of definite descriptions. In these, we single out an object in virtue of its being the unique one satisfying a certain condition; if something is the unique thing satisfying such a condition, we refer to it accordingly. Existence has nothing to do with it. For indefinite descriptions—or even definite descriptions, when no unique thing actually satisfies the defining condition—things are slightly more complex. As we saw in 4.6, the reference is non-deterministic—meaning that factors beyond the semantics fix the reference. Primary amongst these is context, and specifically the intentional acts of the utterer. But none the less, causation need not be required for the appropriate intention, as we will see in a minute.

So much for descriptions. What of proper names? How proper names name is a hard matter. The causal theory of naming has various problems,[11] but let us, for the sake of the present discussion—since the worry is specifically about causation—assume that some version of it is correct. According to this theory, an object is picked out and baptized as *nn* by some agent, *a*. The referent of '*nn*' is picked up by any person, *b*, who talks to *a*, by any person, *c*, who talks to *b*, and so on. Now, causation certainly enters into the transmission of a name's referent; but the causation here is between actual speakers. Supposing that some objects do not exist in no way threatens this. And causation is not required for a baptism—else one could not refer to future objects, which one can. One can, of course, point to a physical object, and so interact with it causally. But one can also pick it out with a description.[12] '*nn*', thereafter, refers rigidly to the object thus selected. Again, non-existent objects in no way threaten this picture. Definite descriptions can be used to pick out non-existent objects just as much as existent ones: 'the object represented by Doyle as living in Baker St, etc., etc.'

Taking intentionality seriously does add an extra dimension to the possibility of baptism, however. As we have seen, picking out an object to name may be performed, not only by a physical act of pointing, but by a mental act of pointing—by simply thinking of the object. Thus, suppose

[11] For a general discussion of the theory, see Devitt and Sterelny (1987), ch. 4.

[12] Kripke (1972), 302.

that there are two people in front of me. By a simple mental act, I can intend one of them rather than the other. We might call this *primitive intentionality*. In this case, there is, of course, a causal interaction between the intender (me) and the intendee. But causation is playing no essential role here: there is exactly the same kind of causal link between me and the non-intended person. There can also be situations where there is no causation of any kind. Thus, for example, I can close my eyes and imagine a scene with two people in it. By an act of pure intention, I can focus on a particular one of these. The intended objects are still, in this case, spatially discriminable (at least in subjective space). But this does not need to be the case either. Suppose, for example, that you tell me about two Ancient Chinese philosophers, Li and Lu. I can intend either of these at will. I may not know anything about them apart from what you have told me. Indeed, you may have told me exactly the same thing about the two (they were philosophers, lived in the Sung dynasty, in the city of Xian). I can still intend whichever I choose. In this case, I know, at least, that one was Li and the other Lu. But, it seems to me, even this minimal amount of distinguishing information may be absent. An act of pure intention can intend an object when there are other indiscriminable objects. How is this possible? That I think, is the nature of the beast. It must be possible, however, because it has actually been done. As I noted in 4.4, the positive and negative square roots of -1, $+i$ and $-i$, are completely indiscriminable in complex arithmetic. (It would make no difference if what we now call '$+i$', we called '$-i$', and vice versa.) But we can intend $+i$ rather than $-i$. Of course, we now have the names to differentiate the two complex numbers; but it was not always thus. At some stage, some mathematician or committee of mathematicians, must have *chosen* one of these objects arbitrarily and called it '$+i$'. Acts of pure intention, it would seem, can be very powerful.

Since intention is a mental act, one might well worry that it falls foul of Wittgenstein's private language argument (*Philosophical Investigations* §§243 ff.); but it does not. In the situation with which Wittgenstein is concerned, a putative act of reference is brought about by fixing on an essentially private object. There is then no public criterion for making a mistake. In such cases, he argues, no act of reference has been performed. Non-existent objects are not, however, private. They are as public as existent objects. And I can focus my attention on one of them, just as much as I can focus my attention on one of a group of people in front of me. In virtue of what I say to you, you can refer to the same thing. It is then perfectly possible for me to make mistakes about which object I originally picked out, which mistakes may be picked up by you. 'Yesterday we were talking about an object you called "Holmes", who lives in Baker St, etc. Now you are telling me that he lives on Olympus, drinks nectar, etc. You've got your wires crossed here: you're talking about Zeus.'

Let me finish this section with a brief discussion of Putnam's 'model-theoretic argument'.[13] Given any theory with a model there are models with different domains that are isomorphic, and so elementarily equivalent, to it. Hence, the theory itself cannot fix what it is about. Putnam uses this as an argument against realism. More recently, Wang (2004) has used it as an argument against noneism. The argument goes as follows. Take any theory, $\mathcal{T}$, the domain of the interpretation of which, according to the noneist, contains some non-existent objects. As long as there is an infinite number of existent things, there will be an isomorphic interpretation of $\mathcal{T}$ the domain of which contains only existing things. Hence, noneism is a hypothesis of which we have no need.

I am in agreement with a number of commentators[14] that the fact that a theory has models that are clearly pathological shows that it takes more than the set of sentences to determine its intended interpretation. And it is clear that the model Wang describes is pathological. One way to see this is to note that one of the sentences in $\mathcal{T}$ is to the effect that some things do not exist: $\mathfrak{S}x\neg Ex$. Hence, to maintain that the correct interpretation of the theory is one in the domain of which all things exist is self-referentially inconsistent. It is not formally inconsistent, of course; for in this interpretation the existence predicate, E, is interpreted so as to apply to only *some* of the existent objects. In other words, it does not have its intended meaning. But to point this out is to give the game away.

This does not, of course, answer the question of what, exactly, it is that makes an interpretation of a theory the correct interpretation. The natural, and, I think, correct answer to this question is that it is the reference relation: the names in the theory must refer to the correct objects.[15] But what determines this? A standard position[16] is to argue that it is some causal connection between the speaker of the language and the

[13] Putnam (1980); page references are to the reprint.

[14] e.g. Devitt (1983), Lewis (1984) (page references are to the reprint).

[15] Putnam (1980), 18, replies to this objection that invoking reference is 'just more theory', and so may itself be reinterpreted. Lewis (1984), 61 f., is right to point out that this is beside the point: the constraint is one that needs to be *satisfied*, not *interpreted as true*.

[16] e.g. Devitt (1983).

object in question, perhaps during the process of baptism involving the name, which determines the referent. This particular answer, as Wang points out, is not available to a noneist. But as we have already observed, even with an object that we do perceive, and so causally interact with, there may be more to the matter than this. Given all the objects in my perceptual field, I can focus my mental attention on just one of them. For example, you can be talking to someone (say, at a party), but really wanting to listen to a conversation that is going on behind your back. In such circumstances, you focus your mental attention on that conversation, though it is not 'perceptually dominant'. Exactly the same can be done with vision. And, again as observed, such attention can single out an object for baptism even in a field of objects that is not brought to your attention causally.

Putnam concedes that his argument can be finessed if one is allowed to appeal to the power of primitive intentionality.[17] But he calls this a 'mysterious faculty of the mind' (Lewis dubs it 'noetic rays'[18]) and complains that it should be rejected by any naturalistic (and sensible) philosopher. I do not see why. If there is a naturalistic account of mental functioning (which I presume there is), then there is a naturalistic account of my undoubted ability to focus my mental attention on a part of, or aspect of, what is phenomenologically present to me. This gives us an account of why it is that the model-theoretic argument fails to work.

7.6 Knowing

The second objection flagged was to the effect that, since we cannot enter into causal connections with non-existent objects, we can know nothing about them. A similar objection is, of course, frequently raised against platonism. It seems to me that any reply to the objection given by a platonist could be adopted—with just as much (or as little) success—by a noneist. But the noneists have other strings to their bow as well.

There is no unique way that one comes to know of the properties of non-existent objects. Depending on the object and the properties in question, there are many ways. For a start, I get to know that the (non-existent) man next door is such that I fear him, by introspection.

[17] (1980), 4. He calls this, unfortunately in the present context, platonism.
[18] (1984), 72.

(I am not suggesting that this is infallible.) I may get to know that the man next door is feared by you by being told. I get to know that Holmes was characterized by Doyle in certain ways by reading the stories.

Another way in which one comes to know some of the properties of non-existent objects is (as Quine suggested about abstract objects) by hypothesis and confirmation. We formulate a theory about how these objects behave and evaluate it according to the normal canons of theory-evaluation, such as simplicity, coherence, adequacy to the (fallible) data, etc. I will give an example of this later in the chapter, so will not pursue the matter further here.

The most distinctively noneist way of coming to know about the properties of a non-existent object is via characterization. An object (existent or otherwise) has those properties attributed to it by the CP, and those that follow from this. We know those properties precisely because we know the CP and can infer from it. Thus, Sherlock Holmes was characterized in a certain way by Doyle. We know that he had those properties, since they are part of the characterization. Further, we know that Holmes had a friend who was a doctor, not because Doyle tells us this, but because he tells us that Watson was Holmes's friend, and Watson was a doctor; we infer the rest. These properties are not properties of Holmes at this world, of course. As we have seen, a characterized object has its characterizing properties at the worlds that realize the way that things are represented to be. In the case of Holmes, this certainly does not include the actual world. In other cases, though, it may.

Mathematical knowledge may also be obtained by characterization. Suppose that we have a mathematical object, c. c is characterized by some mathematical theory, $\mathcal{T}(c)$. Since our grasp of the CP is to explain our knowledge of the facts about c, then $\mathcal{T}$ should, presumably, be something that can *be* grasped. Hence, it is natural to require that the characterization be axiomatic, that is, in effect, that $\mathcal{T}$ be an appropriate set of axioms. Suppose, for example, that $\mathcal{L}$ is the language of arithmetic, formulated in the usual way, with a single constant, 0. Let $\mathcal{T}$ be a set of arithmetic axioms, say the Peano Axioms. Then $\mathcal{T}$ is a set of claims about 0—and various other entities—that characterize its behaviour. Similar comments apply to other mathematical objects and theories.

Do these characterizations obtain at the actual world, or do they, like the Holmes characterization, hold only at other worlds. Nothing so far said forces us to go either way on this issue. However, there would seem to be no particular advantage to supposing that they are true at the actual world. We may therefore treat the cases as alike.[19]

At this point, it is natural to object that this cannot explain our grasp of the properties of mathematical objects, since, in the case of arithmetic, set theory, and similar theories, at least, no axiom system is complete, as we know by Gödel's first incompleteness theorem. Incompleteness per se is not a problem, however. If an axiom system for arithmetic is such that it can prove neither $\psi(0)$ nor $\neg\psi(0)$, this may just show that 0 is an incomplete object: 0 simply fails to satisfy both $\psi(x)$ and $\neg\psi(x)$, just as Sherlock Holmes fails to satisfy both *is left-handed* and *is right-handed* (see 6.4).

However, there is also the stronger version of Gödel's theorem, according to which certain sentences are not only not provable in the axiom system, but can be shown to hold none the less. If this is the case, our grasp of the properties of, say, 0, goes beyond any axiomatic characterization. In answer to this, there are two possible replies.

The first, and obvious, one is that our logic is second-order. As is well known, the second-order characterization of arithmetic is categorical, and so the problem does not arise. There is another possible—and much less orthodox—reply, though. This is to point out that Gödel's first incompleteness theorem claims only that *consistent* (first-order) theories of arithmetic are incomplete. But inconsistent noneist objects are quite possible, so to speak, as I have already observed. It is also well known that there are complete inconsistent theories of arithmetic.[20] Moreover, given that mathematics is a humanly learnable activity, there are arguments to the effect that our arithmetic is both axiomatic and inconsistent. Since these arguments may be found elsewhere, I will not pursue them here.[21] What they show, if correct, is that arithmetic is inconsistent, in which case the problem posed by this version of Gödel's theorem lapses.

One might suggest that our knowledge of the properties of, for example, numbers derives not from characterization, but from Quinean hypothesis-and-confirmation. I do not think that this is the case. The question is what the relevant data is against which the theory is to be tested. In the case of pure mathematics, I do not think there is data independent of our characterization. The case is quite different if we

[19] Note, then, that this makes many claims about both fictional objects and mathematical objects contingent: true at some possible worlds, but not at the actual world—not all though: for example, true identity statements about either kind of object are necessarily true.

[20] See Priest (1997*c*) and (2000*b*).

[21] See Priest (1987), ch. 3.

are testing applied mathematical theories. There we have data about the domain represented. Pure mathematical theories cannot be tested a posteriori. Which brings us to the next objection.

7.7 The a Priori

Objection number three points to the fact that there seems to be a big difference in kind between purely fictional objects and abstract objects, especially mathematical objects. Notably, truths about the former would seem to be a posteriori, whilst truths about the latter would seem to be a priori. What is one to say about this?

The difference in status between mathematical objects and purely fictional objects may be partly explained by the fact that the former are abstract objects and the latter are (normally) concrete. Moreover, as I noted in 7.2, though both are non-existent, the former are necessarily so, whilst the latter are (normally) only contingently so.

But what of the epistemic status of claims concerning the two kinds of objects? For a start, it is not, in fact, the case that all knowledge of abstract objects is a priori. Some is and some is not. For example, we know a priori that the concept *red* (an abstract object) is subsumed by the concept *coloured*. But we do not know a priori that the concept *third planet from the sun* is co-extensional with the concept *planet supporting life*, though this is just as much an (abstract) relation between abstract notions. Nor is it the case that none of our knowledge about fictional objects is a posteriori. It is a priori that Holmes is self-identical.

But concentrate on the sort of examples that people normally have in mind when they make the sort of comparison in question. It would seem that we know a priori that no prime number is the greatest, but not that Holmes lived in Baker St (at least in their respective stories). There certainly appears to be a difference here. But once one looks at the matter more closely, this is not so clear.

The properties of the natural numbers are determined by characterization, say the Peano Axioms. The properties of Holmes are determined, likewise, by characterization—what was written by Doyle. The objects in question have these properties in the worlds realizing the appropriate representations. This is the CP, which is a good candidate for an a priori truth, and is the same in both cases. And it is this that may well be felt to get things wrong. After all, we have to read the Holmes stories to know the properties of Holmes (at his worlds); we do not need to read anything to know about the properties of numbers or sets (at theirs). Or, to put it another way, we are—or, anyway, Doyle was—free to make up the properties of Holmes as he went along. We are not free to make up the properties of numbers as we go along, and neither was anybody else.

Arguably, however, the appearance is misleading. In both cases, we may characterize an object purely by fiat. We know a priori that the object so characterized has those properties (at certain worlds), and this is so whether the characterization is provided by what is told in Doyle's novels, or by the Peano Axioms. Doyle made up the characterization of Holmes by fiat. But the Peano characterization also holds by fiat. Presumably, of course, a fiat that took place a long time ago, and only implicitly—in the practice of counting, adding, and so on; but a fiat none the less.[22]

To see more, it is important to distinguish clearly between two sorts of activity. The first is specifying a characterization; the second is figuring out what follows from it. It is the first of these that we normally think of in connection with fiction (making up a story). It can be done entirely ad lib, and it is this fact that gives fiction its feeling of freedom. But, in certain contexts, we evidently do exactly the same in mathematics. For example, Gödel initiated the study of large-cardinal axioms in set theory. Being a platonist, he assumed that some of these axioms are true and some of them are false, independently of our knowledge. But from a noneist point of view, when we postulate a large cardinal axiom, this is just like extending the Holmes stories (see 6.3). And there is no right or wrong way to extend the characterization of sets, any more than there is a right or wrong way to tell a new Holmes story: any way will do (or at least, any way that is compatible with what has gone before).

The second sort of activity, the drawing out of consequences, is what we normally think of first in connection with mathematics. The characterizations of mathematical objects are normally now fixed: mathematics comprises the deduction of what follows from these. There is nothing a posteriori about this: the consequences are governed by the laws of logic. It is this that gives mathematics its a priori feeling. But it is clear that we engage in the second sort of action with respect to fiction as well. When we come out of a film, we argue about the characters, inferring

[22] In some branches of mathematics one gets to know the characterization explicitly. For example, one is normally given the axioms of group theory in the first lecture on the topic. But with numbers it is not (normally) like this. One absorbs the Peano Axioms implicitly when one learns to count, add, etc.

from what was shown or said. And the phenomenology of this process is, in fact, very similar to arguing about mathematical objects, though the predicates concerned in arguing about fictional objects are mostly vague, and so interesting cases are rarely cut and dried in the same way that they are in mathematics.

There is at least one further point of dissimilarity that we may observe. Standardly, as I noted in 6.2, not all the representation in a work of fiction is explicit. Thus, though Doyle never tells us (or could have told us) this, it is part of the representation in the Holmes stories that there are no aeroplanes. This is a clear-cut case. But there will be cases that are not so clear. In science-fiction films and stories, for example, it is not always clear what the author intends us to carry over about the laws of nature from real life into the fiction. When we argue about works of fiction, therefore, part of what we may be arguing about is what, exactly, the representation is. This is not the case in mathematics—at least, modern mathematics—where nothing beyond the axioms may be appealed to.

Let me summarize what has been learned in the preceding discussion. There are some important differences between paradigm fictional and mathematical objects, especially concerning the modal status of their existence. There may also be some differences when it comes to a priori and a posteriori knowledge about them; but not substantial differences of the kind one might have thought.

7.8 Applying Mathematics

Let us turn to the fourth objection. How can non-existent objects tell us anything about existent ones? Routley (2003) gestured at a noneist solution to this problem. Facts about non-existent objects can inform us about existent objects since the facts about actual objects may *approximate* those about non-existent objects. Think, for example, of a frictionless plane, an ideal, but non-existent, object. A real plane is not frictionless, but it can be approximately frictionless. Hence, with suitable provisos, if A is true of the ideal plane, A is approximately true of the real plane. Thus, if A is a claim to the effect that an object slides a certain distance across the ideal plane in time t, we can infer that an object will slide the same distance across the real plane in a time $t \pm \varepsilon$, where ε is a contextually determinable real number.

Even if something like this is right, the answer can be only a partial one. For on many occasions we use numbers, non-existent objects, to tell us *exactly* how an existent object will behave. Thus, for example, suppose there is a particular particle, say an electron. Suppose that it is moving with a constant velocity **v**, and that it moves for a time **t**, through a distance **d**. Here, **v**, **t**, and **d** are particular physical, not mathematical, quantities. But each of them can be assigned a certain numerical magnitude, v, t, and d, respectively, by some measuring procedure (using clocks, rulers, etc.). Thus, for example, there is some family of observable properties, P_n, of the distance such that:

(⋆) $P_n\mathbf{d}$ iff $d = n$

This establishes a correlation of a certain kind between **d** and d. Call biconditionals of this kind *bridge laws*. Now, a law of motion tells us that:

$$d = v \times t$$

Thus, if we establish by observation, via the bridge laws, that $v = 3$ and $t = 6$, we infer that $d = 3 \times 6 = 18$, and so that $P_{18}\mathbf{d}$. We have used pure mathematical facts to infer something about a physical quantity. Nor are we dealing with ideal objects here; the particle in question is a real-life particle.

How, then, is one to explain the fact that properties of non-existent objects can tell us something about existent objects? Actually, exactly the same question can be posed for platonism, and the answer in both cases is the same. The physical quantities in question have certain properties, and the mathematical quantities have other properties. But we can move between the one and the other because these properties have the same structure, and, specifically, because the correlation established by the bridge laws is an isomorphism. Since mathematical objects may not have their relevant properties at the actual world, we have to understand the bridge laws in a particular way. Thus (⋆), for example, has to be understood as:

(⋆⋆) $P_n\mathbf{d}$ iff $w \Vdash^{+} d = n$

where w is any world that realizes the truths of arithmetic. But the bridge laws still fulfil the function of allowing us to move back and forth between the properties of the physical quantities and those of numbers.

This sort of explanation is quite general. A science, or a branch of it, concerns certain physical quantities, $\mathbf{q}_1, \ldots, \mathbf{q}_m$. These have associated numerical magnitudes, $q_1, \ldots, q_m$, determined by bridge principles of the kind (⋆⋆).[23] In virtue of certain physical states of affairs and the bridge principles, we have some mathematical relation, $F(q_1, \ldots, q_m)$—typically in physics, this would be a differential equation—and working with this we can establish various facts about the q_i, and hence, via the bridge principles, certain physical states of affairs.[24]

Thus, we can use facts about mathematical objects to infer facts about physical states precisely because the two have the same structure. That a certain relation obtains between the mathematical objects can be determined a priori from their characterizations; but which physical relations are isomorphic to which mathematical relations is an a posteriori fact. Its discovery is that of a law of nature. This explanation, which depends simply on there being certain correlations between properties of physical magnitudes and properties of mathematical magnitudes, in no way depends on the numerical magnitudes being existent. All it depends upon is their having the right *Sosein* at the appropriate worlds.

Before we leave the question of applying mathematics, let us return to worlds and their properties. The semantics of Part 1 of the book is a mathematical structure, couched within set theory. But it can be applied to tell us something about worlds, which are not. It is applied in exactly the way that I have just described, though what the mathematical theory of worlds represents is no longer physical reality, but certain non-existent objects. If the relevant aspects of the mathematical semantics are isomorphic to those of worlds, that is, if the representation gets things right, then we may infer facts about worlds from those about their set-theoretic representations via appropriate bridge principles. How do we know whether the set theoretic representation gets things right? In the same way that we test any applied mathematical theory. There are certainly other possible semantics for an intentional language. We determine which it is the most rational to accept in terms of the usual criteria of theory acceptance, such as simplicity, adequacy to the data, and so on. What counts as the data in this case? The sorts of claims that we are

[23] The properties, P_n, employed need not all be observable. Some may be establishable only by inference.

[24] If one is not a realist about space and time (which I am), one may suppose that there are no actual quantities of space and time, but that talking of such is just a way of talking about certain relationships between objects in space and time. One might therefore object to the particular example I used above. If one does, however, a general account of the same form can still be given. The physical quantities in question are just different (depending on how, exactly, talk of space and time is cashed out).

inclined to make using intentional notions and the inferences that we are inclined to draw concerning them. Such data is, as always, fallible, and may be rejected in the light of overall coherence. None the less, it would be irrational to accept an account of intentionality that ruled out most or much of what we take to be the case concerning intentionality—on that very count.

But why, it might be asked, do we not simply characterize the worlds in question in the appropriate fashion, and infer their properties from characterization? The answer is as follows. One can, indeed, characterize worlds and their properties in any way that one wishes. The worlds, so characterized, have their properties at the appropriate worlds, but these may not be actual. In other words, the statements attributing those properties may not actually be true. But when it comes to semantics, we are after, not just a story, but the truth.

7.9 Platonism

Let us move, finally, to the last objection. This is to the effect that noneism is just platonism in disguise. According to this objection, the translation manual at Table 7.1 shows that a noneist is simply a platonist with an unusual vocabulary.[25]

The objection might well be reinforced by the fact that, in answer to some of the previous objections, the noneist and the platonist can say much the same thing. Note, of course, that for the reduction to be a general one, it must be made with respect, not just to abstract objects, but to all non-existent objects—purely fictional objects, worlds, and so on.

There are many things to be said about this objection. The first is that translation manuals are symmetric. Hence, to suppose that the manual establishes that noneism is reducible to platonism is quite question-begging—at least without further argument. We might just as well say that platonism reduces to noneism. Without such considerations,

[25] An objection to the effect that Routley's view collapses under this translation is made in Lewis (1990). Burgess and Rosen (1997), 224, dismiss Routley's view summarily with an appeal to Lewis's paper. To the extent that they have reasons of their own (p. 188 f.), they perpetuate the confusion that noneists appeal to some kind of being other than existence (see 5.2 and 5.3). They then say that it does not help to replace 'there is' with 'for some': it is not easy, they claim, to understand what the difference is between 'exists' and 'some'. They could simply have reflected on the sentence 'I thought of something I would like to give you as a Christmas present, but I couldn't get it for you because it doesn't exist.'

Table 7.1

Noneist	**Platonist**
is an object	exists
exists	is a concrete object

we can just as well say that Plato was a noneist as that Routley was a platonist.[26]

Next, there are, in any case, differences between the two positions. Crucially, the noneist subscribes to the CP; the platonist, at least as usually understood, does not. The noneist claims that any instance of the CP characterizes a perfectly good (though maybe non-existent) object. The platonist does not normally say that an arbitrary characterization characterizes an existent object. Numbers, sets, geometrical lines and points, all these exist. But there is no reason to suppose that any old axiom system—or story—specifies existent objects.

There is a version of platonism that does claim this, however: *plenitudinous platonism*.[27] The plenitudinous platonist holds exactly that there is nothing privileged about the axiom systems for numbers, geometric objects, etc. Every axiom system characterizes equally good abstract objects. The thought that every (consistent) axiom system has a model gives some credence to plenitudinous platonism. The fact that a sentence has a model does not show that it is really satisfiable by certain objects. For example, one can construct a model of the first-order existence '$\exists x(x$ is married $\wedge\, x$ is a bachelor)', though there is no existent object, x, such that x is married $\wedge\, x$ is a bachelor. Still, models are very much like realities, and the fact that every (consistent) characterization has a model at least gives us a model (so to speak), of what it would be like for every characterization to characterize existent objects (from a platonist point of view).

The confluence between noneism and plenitudinous platonism is still not right, though. A thoroughgoing noneist holds that every characterization characterizes an object. And here, 'every' means *every*.

[26] Actually, this is one place where the translation manual is certainly not adequate; for Plato held that the forms were not only existent (real), but that they were *more* existent (real) than concrete objects. No noneist has ever claimed that abstract objects are more existent objects than concrete ones.

[27] I take the name from Field (1998). The view is advocated by Balaguer (1995), where it is called 'full blooded platonism'. Balaguer defends this platonism against the epistemological objection of 7.6 on grounds very similar to those employing characterization that I used there.

Even inconsistent characterizations do this. This diet is probably too rich for even a plenitudinous platonist. Platonists are characteristically very much attached to consistency. So this is an important difference between the noneist and the plenitudinous platonist. Of course, there is still another position out there. This belongs to what we might call the *paraconsistent plenitudinous platonist*. This is a platonist who has foresworn classical logic, and is prepared to endorse a paraconsistent logic.[28] Such a platonist can hold, quite generally, that every characterization characterizes an existent object.

Of course, this sort of platonist cannot hold that every object characterized has its characterizing properties at this world. As we saw in 4.2, if the CP is true at this world, the world is trivial. Hence, the paraconsistent plenitudinous platonist must hold that many of the objects characterized by the CP have their characterizing properties at other worlds. They exist, however, at all worlds.[29]

At this point, the differences between noneism and platonism are disappearing fast. And it must be said that it is the platonist who is making all the concessions. This is a reason to say that the sort of platonism that is left is really noneism in disguise, and not vice versa.

But, since the matter of different possible worlds has now arisen, there is, in any case, still a point where the translation manual breaks down—in modal contexts, and specifically in claims concerning modal status. Thus, consider the claim that Holmes does not exist, but could have done so. This is a claim to which the noneist will assent. The translation is that Holmes is not a concrete object, but could have been. This hardly seems to be true. If Holmes is not a concrete object, what is he? He is not a set, number, property, or other sort of abstract object. And if he is, since abstract objects have their modal status necessarily, it is not possible for him to be a concrete object.

Conversely, Routley did exist, but might not have done so (had his parents not met, for example). The translation of this is that Routley was

[28] This position is mooted in Beall (1999).

[29] A difference between standard platonism and the noneism of the kind explained in this chapter is that, typically, platonists tend to say that the familiar claims about mathematical objects are actually true; this is not the case for noneism of the kind explained. The difference is superficial, however. A noneist could, without too much change to what I have said, hold that standard mathematical objects have their characterizing properties at the actual world; and conversely, as we have just seen, a platonist could hold that mathematical objects have their familiar properties at worlds other than the actual.

a concrete object, but might not have been. In another possible world, he was a set? Of course, there are worlds where Routley is a set—merely consider the characterizing condition: $x = x$ and Routley is a set. This, like all characterizing conditions, is realized at some worlds; but they are not possible: concrete objects cannot be abstract objects.

A similar point can be made in terms of the explicit characterizations of abstract and concrete objects of 7.2. The number 3 is an abstract object. On the understanding of 7.2, this means that *if 3 did exist, it would not causally interact with us*. Under the translation manual, then means: *if 3 were concrete, it would not causally interact with us*. But this is false: had 3 been a concrete object, then we would have been able to interact causally with it. Of course, a platonist might try to fashion some other criterion for being an abstract object; but we have seen that such tend to be problematic. And in any case, it remains the fact that the truth of this counter-factual is still not preserved under translation.

There may well be other statements whose truth-value is not preserved under translation. But we have seen enough. Even the attenuated form of platonism, paraconsistent plenitudinous platonism, is still distinct from noneism.

7.10 Conclusion

In this chapter, we have looked at mathematical objects and worlds as non-existent objects. Leaning on what was said in previous chapters, and particularly the account of characterization, we have seen how a natural account of these objects, their properties, and our abilities to refer to and know about them, can be given—an account which is not subject to some natural objections. In fact, in the last three chapters, we have taken in most of the standard objections to noneism. There is one further objection, however. It is not a standard one, but—for my money, at least—it is the hardest. This is the subject of the next, and last, chapter.

8
Multiple Denotation

8.1 Introduction: A Paradox of Denotation

In this chapter I will explain and give a solution to what I take to be the hardest problem for noneism—though for reasons that will become clear, it may not strike many people that way. This concerns a paradox in the family of paradoxes of self-reference, and specifically in the family of denotation paradoxes, such as Berry's. Normal paradoxes of this family are solved very happily by allowing contradictions to occur at the base world, @. In this case, the logic is paraconsistent, and the failure of Explosion ($A, \neg A \vdash B$) ensures that the contradictions generated by the paradoxes occur simply as isolated singularities. What distinguishes the paradox that we will be looking at here is that it cannot be handled in such a way—at least, not in the most obvious fashion.

I will first explain the paradox. We will next look at some possible solutions that are not satisfactory. Then I will explain the one that strikes me as the most plausible in this context. This will require (perhaps unsurprisingly) revising the way that denotation is taken to work. As we will see, the major consequences of the revision concern identity. The solution is demonstrably adequate, in the following sense. One can show that the theory which generates the paradox, provided that it is based on the treatment of denotation advocated, though inconsistent, is non-trivial. The proof of this fact, together with those of some other technical claims, is given in the technical appendix of this chapter.

8.2 Semantic Paradoxes of Self-Reference

The semantic paradoxes of self-reference are generated by the naive principles that govern our semantic notions, and especially those of truth, satisfaction, and denotation. Let $[\cdot]$ be an appropriate name-forming functor, and let us write T for the truth predicate, S for the (one place) satisfaction relation, and Δ for the denotation relation. Then these principles are, respectively:

Truth: $T[A] \dashv\vdash A$, where A is any (closed) sentence

Satisfaction: $S(x, [A(y)]) \dashv\vdash A(x)$, where x is free on substitution

Denotation: $\Delta([t], x) \dashv\vdash t = x$, where t is any (closed) term.

The exact nature of the connection between the left- and right-hand sides of these principles is an interesting question, though it is not crucial for the present discussion. The minimal connection of bi-deducibility will do here.

It is well known that these principles, when combined with self-reference and a few simple principles of logical inference give rise to contradiction.[1] What is less well appreciated is that the denotation paradoxes are distinctive in that they, unlike the paradoxes of truth and satisfaction, require descriptions of some kind to generate the paradox.[2] The paradox we will be concerned with is a paradox of this kind. Though a definite description operator will do just as well, the simplest procedure is to employ an indefinite description operator. I remind (from 4.7) that such an operator satisfies the condition:

Des: $\mathfrak{S}xA(x) \vdash A(\varepsilon xA(x))$

provided that $\varepsilon xA(x)$ is free when substituted for x in $A(x)$.

8.3 The Paradox of Hilbert and Bernays

The paradox appeared first (as far as I know) in Hilbert and Bernays' *Grundlagen der Mathematik*,[3] where it is used to show that a consistent theory cannot contain its own denotation function (in the same way that the Liar paradox is deployed in Tarski's Theorem to show that such a theory cannot contain its own truth predicate). Basically, the argument is very simple. Suppose that we are talking about terms that refer to numbers,

[1] See e.g. Priest (1987), ch. 1; (1995*a*), ch. 10.

[2] The theme is discussed further in Priest (forthcoming *a*).

[3] Hilbert and Bernays (1939), 263–78.

and consider the term 'the denotation of this term plus one'. This denotes some number, n, but then it denotes $n+1$ too. So $n=n+1$, and $0=1$.

Let me generalize the argument and make it more precise.[4] Let us suppose that the language contains that of first-order arithmetic, so that we may use its self-referential powers. We suppose that we have an appropriate Gödel coding, #, so that if t is any term of the language, $\#t$ is its code, and $[t]$ is the numeral of $\#t$. Employing this, it is straightforward to show that, if $t(x)$ is any term, there is a closed term, t', such that:

SR $t' = t([t'])$

t' is 't of this very term'.[5]

A formalization of the paradox now goes as follows. Let f be any one-place function symbol. We know from **SR** that there is a closed term, t, such that:

1. $t = f\varepsilon y\Delta([t], y)$

I note the principles of inference involved at each of the next steps. SI is the substitutivity of identicals.

$t = t$	Identity
$\Delta([t], t)$	**Denotation**
$\mathfrak{S}y\Delta([t], y)$	Generalization
$\Delta([t], \varepsilon y\Delta([t], y))$	**Des**
$t = \varepsilon y\Delta([t], y)$	**Denotation**
$t = ft$	By 1 and SI

To do damage, we now let f be the successor function. Then we have a t such that $t = t+1$, and hence, $0 = 1$.

4 This follows Priest (1997*b*).

5 Strictly, an extra assumption is needed, that the diagonalization function is represented in the language by an appropriate *term*. If $r(x)$ is any term with one free variable, x, the diagonalization of $r(x)$ is $r([r(x)])$. We assume that there is a term, $g(x)$, such that if m is the (code of) the term (with code) n then it can be proved that $g(\mathbf{n}) = \mathbf{m}$. (Here, boldfacing represents the appropriate numeral.) The proof then goes as follows. Consider the term $t(g(x))$. Call this $r(x)$. Its diagonalization is $t(g([r(x)])$. Call this t'. Since the diagonalization of r is t', $g([r(x)]) = [t']$. Hence, $t(g([r(x)])) = t([t'])$. But the left-hand side is exactly t'.

8.4 Solutions

So much for the problem. What of solutions? And what, anyway, has this to do with noneism?

Probably the first thought that will occur to most people—certainly those who wish to maintain a consistent account of the paradoxes of self-reference—is that the Denotation principle cannot be accepted in full generality. Thus, if we adopt various solutions to the semantic paradoxes, only some instances of the naive principles governing our semantic notions are accepted. For example, if we endorse a Tarskian solution to the paradoxes, the T-schema and its relatives hold only in restricted forms. The inadequacies of a Tarskian solution are well known, though. Indeed, there are good reasons for supposing that the schema and its like ought to hold in full generality. If one supposes that one of the functions of the truth predicate is as an inverse to quotation—for a deflationist about truth, this is its only function—any restriction on the T-schema will seem entirely unmotivated. Of course, there are (consistent) solutions to the paradoxes that maintain the integrity of the naive semantic principles. If they do this, they then have to reject certain principles of inference. The favourite candidate is the Law of Excluded Middle. In the paradox at hand, though, the rejection of the Law will be of no avail. The argument of the previous section does not employ it. Indeed, its logical debts are pretty minimal.

In any case, and setting these issues aside, there are general reasons why, it seems to me, no consistent solution to the semantic paradoxes is adequate. This is not the place to discuss them,[6] but let me just say that I take it that any adequate solution to the semantic paradoxes of self-reference must allow contradictions to occur, but employ a paraconsistent logic to isolate them.

That move will not help in this context though. The paradox deduces not a simple contradiction, but that $0 = 1$. This certainly does contradict the fact that $0 \neq 1$. But it is much worse than that. For given that $0 = 1$, pretty much anything can be deduced in arithmetic. The sentence is triviality-producing.

It might be suggested that what needs to be rejected is some arithmetic principle involved. Thus, we have deduced that $t = t + 1$, but we need

[6] The matter is discussed at great length in Priest (1987) and (1995*a*).

various principles concerning subtraction to conclude that $0 = 1$. This particular contradiction can, indeed, by handled in this way. There are inconsistent arithmetics where there are truths of the form $n = n + 1$, but where $0 = 1$ does not hold.[7] But even this will not help. For applying the argument to the successor function is just one way of generating trouble. A more direct way is applying it to the zero function, ζ. Thus, let $\zeta(x) = 1$ if $x = 0$, and 0 if $x > 0$. Then the argument delivers a term, t', such that $t' = \zeta(t')$. But then if $t' = 0$ then $t' = \zeta(t') = 1$; so $0 = 1$. And if $t' > 0$, $t' = \zeta(t') = 0$; so $0 = 1$, as before.

Another way of trying to solve the problem is by appealing to denotation failure. Thus, if t is equal to n, it is equal to $n + 1$ too. This would seem to be a good reason to infer that 't' has no denotation. Standard logic assumes that all terms denote, but this assumption can be rejected in favour of some free logic. How this blocks the argument depends on exactly how the free logic is implemented. Generalization is a natural casualty, but so may Identity be. **Denotation** may also be restricted to those terms that denote.

Unfortunately, this strategy is not as straightforward as it seems either.[8] For even if a description is not guaranteed a denotation, we can use it to define one that is, thus: $\varepsilon x((\mathfrak{S}yA(y) \land A(x)) \lor (\neg\mathfrak{S}yA(y) \land x = 0))$. The argument then proceeds much as before. This time, at least, though, it is fallacious from a paraconsistent perspective, since it employs the Disjunctive Syllogism. One can, indeed, show that, assuming that terms may fail to denote, a theory endorsing all the relevant principles, but based on a paraconsistent logic is inconsistent but non-trivial.[9]

But now, and to come at last to noneism. Even this option is not open to a noneist. Since one can think of an object as specified in any way one likes, then all terms must denote—and, for good measure, Generalization is guaranteed too. There seems to be very little room to manœuvre. This is the hard problem for noneism.

8.5 Multiple Denotation Semantics

I noted that if 't' denotes some number, n, it denotes its successor. We might infer from this that 't' has no denotation. But we could infer, with

[7] The matter is discussed further in Priest (1997*c*).
[8] For details of the following, see Priest (1997*b*).
[9] See Priest (1999).

equal justification, that '*t*' has more than one denotation. This is the idea that the rest of this chapter will explore.

In the context of paradox, it is, in fact, a natural strategy. In the case of more usual paradoxes of self-reference, the arguments concerned conflict with the assumption that a sentence must be either true or false, but not both. Some try to solve these paradoxes by rejecting the thought that a sentence must be true or false. That is, they reject the Law of Excluded Middle. Leaving aside the adequacy of this, dialetheism plays the opposite side of the street. Every sentence has at least one truth value; it may just have more than one. The same choices are possible in the case of denotation. The paradox comes into conflict with the assumption that every term has one and only one denotation. We can suppose that some terms have no denotation, and so move to a free logic. The adequacy of this move I discussed in the previous section. Alternatively, we can assume that a term can have more than one denotation. This is the analogue of the dialetheic move. Thus, just as self-reference and the naive semantic principles force on us the fact that, perhaps despite our best intentions, a sentence may have more than one truth value, so they force on us the fact that a term may have more than one denotation.

What is required to handle this situation is a logic of multiple denotation. In this section, I will give the semantics for such a logic. We will then go on to apply it to the paradox argument.[10]

As is clear from the paradoxical argument, the machinery of intentionality, and of worlds in general, is quite irrelevant. The paradox arises with just extensional connectives and quantifiers, descriptions and a denotation predicate. In what follows, I will therefore simplify by ignoring all the complexities that arise due to worlds.[11] In particular, $\Vdash_s^{\pm}$ is a relationship between an interpretation (mention of which is normally omitted) and a formula. The world parameter on its left disappears.

An interpretation is constituted by a structure $\langle D, \delta, \varphi \rangle$, where D is the domain of quantification, and δ assigns a denotation in D to each constant, a function to each function symbol, and an extension and co-extension to every predicate, in the usual way. Identity has its standard extension, and we require the extension and co-extension of each predicate to be exhaustive, but not necessarily exclusive.[12]

[10] Semantics of this kind were developed in a quite different context in Priest (1995*b*).

[11] Extending the construction to include the intentional machinery is a non-trivial matter.

[12] This gives the paraconsistent logic *LP* (see Priest 1987: ch. 5.), which validates the Law of Excluded Middle (LEM). We could allow extension and co-extension not to be exhaustive as

The only essential difference in the multiple-denotation case is that the choice function, φ, is such that for every τ, φ_τ picks out a subset of its argument, not a member. More precisely, if $X \subseteq D$ then:

$\varphi_\tau(X) \subseteq X$

if $X \neq \phi$, $\varphi_\tau(X) \neq \phi$

The denotation of each term, t, of the language, $\delta_s(t)$, is no longer a single term, but a set of terms. Thus:

If x is a variable, $\delta_s(x) = \{s(x)\}$

If c is a constant, $\delta_s(c) = \{\delta(c)\}$

If f is an n-place function symbol, $\delta_s(ft_1 \ldots t_n) = \{\delta(f)(a_1, \ldots, a_n) : a_1 \in \delta_s(t_1), \ldots, a_n \in \delta_s(t_n)\}$

The natural generalization of the denotation conditions of 4.5 to multiple denotation (ignoring the complexity added by worlds) is as follows. Let τ be $\varepsilon x A(x)$. Then:

$\delta_s(\tau) = \varphi_{\overline{\tau}}\{d\colon \Vdash^+_{s(x/d)} A(x)\}$	if this set is non-empty
$\varphi_{\overline{\tau}}(D)$	otherwise

This would, in fact, provide what is required. But if $\varepsilon x A(x)$ can have multiple denotations, even if some things do satisfy $A(x)$, why should it denote just some of those? Thus consider the description 'a thing that is red'. This denotes some red things; but why should it not also denote some non-existent things as well: maybe things that are red at some other worlds? Thus, if τ is $\varepsilon x A(x)$, we will let:

$$\delta_s(\tau) = \varphi_{\overline{\tau}}\{d\colon \Vdash^+_{s(x/d)} A(x)\} \cup \varphi_{\overline{\tau}}(D)$$

Note that every term has a non-empty denotation. But note also that for constants, variables, and, more generally, all terms that do not contain a description, this is a singleton; such terms, then, effectively, still have a single denotation. This could be changed, but multiple-denotation for descriptions will suffice for our needs.[13]

well. This would invalidate the LEM. However, in the present context, having the Law is an advantage, since the non-triviality argument of the appendix then shows that all variants of the paradox fail even in its presence.

[13] Allowing for constants to have multiple denotations introduces no major differences. But allowing for variables to have multiple denotation does, since this affects the way that quantifiers function.

Turning from denotation conditions to truth/falsity conditions: for atomic formulas, $Pt_1 \ldots t_n$:

$$\Vdash_s^{\pm} Pt_1 \ldots t_n \quad \text{iff for some } x_1 \in \delta_s(t_1), \ldots, \\ x_n \in \delta_s(t_n), \langle x_1, \ldots, x_n \rangle \in \delta^{\pm}(P)$$

Thus, atomic sentences are true/false (given an interpretation of the free variables) iff *some* denotations of its terms are in the extension/co-extension of the predicate.[14] As is clear, these conditions reduce to the usual ones when the denotation of a term is a singleton, as it is for terms not containing an ε. Note that even if $\delta^+(P)$ and $\delta^-(P)$ are disjoint, $Pt_1 \ldots t_n$ may well be both true and false, since some denotations of the terms may satisfy P and some may satisfy its negation. There could therefore be truth value gluts, even if all the predicates were classical (that is, had exclusive and exhaustive extension and co-extension).

There is another option, which is to go for *all* instead of *some*:

$$\Vdash_s^{\pm} Pt_1 \ldots t_n \quad \text{iff for all } x_1 \in \delta_s(t_1), \ldots, \\ x_n \in \delta_s(t_n), \langle x_1, \ldots, x_n \rangle \in \delta^{\pm}(P)$$

This option generates truth value gaps rather than gluts. I will comment on it further later.

The truth/falsity conditions for propositional connectives and quantifiers are the same as before; and validity is defined in terms of truth-preservation in all interpretations.

8.6 Properties of the Semantics

Because the only thing that has really changed in these semantics is the behaviour of descriptions, the rest of the logic remains unchanged (except that, as we noted, the semantics will require the possibility of contradictions holding). What of descriptive terms? Proofs of all the points made in what follows can be found in the technical appendix to the chapter.

The relationship between variables and the denotations of terms is a little tricker in this case than in the usual case. However, one can establish the appropriate facts and so show that universal instantiation is valid: $\mathfrak{A}xA(x) \models A(t)$, provided that t is free when substituted for x. There is

[14] If P is a propositional parameter, so that there are no t_is, then 'for some $x_1 \in \delta_s(t_1), \ldots, x_n \in \delta_s(t_n), \langle x_1, \ldots, x_n \rangle$' is to be understood simply as '$\langle\rangle$'.

also an extra restriction in this case, namely that t is not itself in the scope of an ε in $A(t)$. **Des** is also valid: $\mathfrak{S}xA(x) \models A(\varepsilon xA(x))$, provided that $\varepsilon xA(x)$ is free when substituted for x, and, again, that $\varepsilon xA(x)$ is not in the scope of an ε in $A(\varepsilon xA(x))$—which it is not in the paradox argument.[15]

Though Universal Instantiation is valid, subject to the restrictions, Particular Generalization fails. That is, $A(t) \not\models \mathfrak{S}xA(x)$ (even when t is free in $A(t)$ and not within the scope of an ε). To see this, merely consider an interpretation where the domain has two distinct objects, a and b. Let t be a term such that $\delta(t) = \{a, b\}$, e.g. $\varepsilon x\, x = x$. Let the extension of P be $\{a\}$ and the extension of Q be $\{b\}$. Then it is easy to check that $\Vdash_s^+ Pt \wedge Qt$. However, there is no d in the domain such that $\Vdash_{s(x/d)}^+ Px \wedge Qx$, since $a \in \delta^+(P)$, but $a \notin \delta^+(Q)$; and vice versa for b. Hence, $\nVdash_s^+ \mathfrak{S}x(Px \wedge Qx)$.[16] However, one may show that restricted versions of it are valid. In particular, provided that x has only the occurrence indicated in the atomic context $Pt_1 \dots x \dots t_n$ then $Pt_1 \dots t \dots t_n \models \mathfrak{S}xPt_1 \dots x \dots t_n$.

8.7 The Paradox Revisited

What of identity? This is clearly reflexive: let $a \in \delta_s(t)$; then $\langle a, a\rangle \in \delta^+(=)$, so $\models t = t$. (Note, though, that if t has multiple denotations, $t = t$ is also false.) It is also symmetric, $t_1 = t_2 \models t_2 = t_1$, since if $\langle a, b\rangle \in \delta^+(=)$, then $\langle b, a\rangle \in \delta^+(=)$. It is not, however, transitive: $t_1 = t_2, t_2 = t_3 \not\models t_1 = t_3$. To see this, arrange an interpretation where $\delta_s(t_1) = \{a\}$, $\delta_s(t_2) = \{a, b\}$, $\delta_s(t_3) = \{b\}$ (and a and b are distinct). Then $\Vdash_s^+ t_1 = t_2$, since a is in the denotation of both terms; similarly, $\Vdash_s^+ t_2 = t_3$, since b is in the denotation of both terms. But nothing is in common between the denotations of t_1 and t_3; hence, $\nVdash_s^+ t_1 = t_3$. Transitivity of identity is a special case of the substitutivity of identicals. Hence, this must fail too. Here is another counter-example: $t_1 = t_2, Pt_1 \not\models Pt_2$. For a countermodel to this, choose an interpretation where $\delta_s(t_1) = \{a, b\}$, $\delta_s(t_2) = \{b\}$, and

[15] As to why these inferences fail when the substitution is into the scope of an ε, see 8.10, Lemma 19.

[16] In virtue of this, one might wonder where the standard proof of the equivalence between Universal Instantiation and Particular Generalization breaks down. The answer is that contraposition is not valid in the semantics. Thus, for any term, t, $\models t = t$. Hence, where P is a propositional parameter, $P \models t = t$. But $\neg t = t \not\models \neg P$. To see this, just choose an interpretation at which P is true but not false. Then the conclusion is false, but not true. Now, let t have two denotations, a and b. Then $t = t$, and so its negation, is both true and false.

$\delta^+(P) = \{a\}$. It is not difficult to check that the premises hold, but the conclusion does not.[17]

Since substitutivity fails, it may fairly be asked whether this provides a solution to the Hooded Man paradox of Ch. 2. The semantics certainly invalidates substitutivity, but only where the terms involved have multiple denotations. In the argument of 2.5 the terms in question are not descriptions but constants, which have single denotations according to the preceding semantics. Of course, we could modify things to allow for constants to have multiple denotations too, and so for substitutivity to fail for these. But even so, this would not appear to be the right way to handle the paradox: it does not seem to be a case of multiple denotation at all. In the Hooded Man paradox, 'Nescio' and 'Cain' refer to one and the same person, and, as far as the example goes, to nothing else.

We can now return to the paradox of denotation. As may be seen by checking the deduction in 8.3, and assuming the correctness of **Denotation**, all the steps prior to the last are ones which are valid.[18] The last, however, is an instance of the substitutivity of identicals, and so fails. More than this, one can show that no argument of the kind in question is possible in this logic: there is an interpretation that renders **Denotation** true, but not $0 = 1$ or any other sentence false in the standard model of arithmetic (even when the restriction on Universal Instantiation and **Des** concerning substitution within ε-terms is lifted). The proof of this non-triviality result is a persistence-type argument. I give it in the technical appendix to the chapter. Hence, the paradox of multiple denotation is solved; and since every term has a denotation, noneist principles have not been compromised.

What happens if we take the alternative *all* truth/falsity conditions for multiple denotation? Substitutivity now holds. In fact, $t_1 = t_2$ is not true if either of these terms has multiple denotations. Hence, if $t_1 = t_2$ is true, the terms have the same singleton denotation, and substitutivity holds. But now the argument breaks down at other points. In particular, Identity fails, since if t has multiple denotations, $t = t$ is not true, since not all

[17] Non-transitive, non-substitutional, notions of identity arise in other contexts too. Thus, fuzzy identity is like this (see Priest 1998). When identity is given a natural second-order definition, it also behaves like this (see Priest forthcoming *c*). It would be very natural to apply the second of these notions of identity in the case at hand, and so solve the paradox of denotation in this way as well.

[18] This includes the proof of **SR** given in n. 5. Althought this uses SI, the substituted terms are not descriptions, so the instance of SI employed is legitimate.

its denotations are identical. For similar reasons, $\mathfrak{S}x\, x = t$ fails if t has multiple denotations. Thus, given **Denotation**, $\mathfrak{S}x\Delta(\langle t\rangle, x)$ also fails. One cannot, therefore, truly say that t denotes something—which it does in this case. In other words, the denotation predicate is not faithful to real denotation. This makes this construction, as a solution to the paradox, implausible. For in the case of multiple denotations, all terms *do* denote. Note that, in the *some* approach, $\mathfrak{S}x\, x = t$ is a logical truth. (Let $a \in \delta_s(t)$. Then $\Vdash^+_{s(x/a)} x = t$, and hence $\Vdash^+_s \mathfrak{S}x\, x = t$.) Hence, given **Denotation**, $\mathfrak{S}x\Delta(\langle t\rangle, x)$ holds.

8.8 Definite Descriptions

The paradoxical argument that we have been dealing with fails because the crucial term involved has multiple descriptions. A natural thought is that this is a function of the fact that we have used an indefinite description: if we were to use a definite description, denotations would be unique, and so the argument would go through. Is this so? Using definite descriptions, we have a closed term, t, such that:

1. $t = f\iota y\Delta([t], y)$

The natural argument now goes as follows:

$t = t$	Identity
$\Delta([t], t)$	**Denotation**
$\neg\mathfrak{S}x(\Delta([t], x) \wedge t \neq x)$	(⋆)
$\Delta([t], t) \wedge \neg\mathfrak{S}x(\Delta([t], x) \wedge t \neq x)$	Adjunction
$\mathfrak{S}y(\Delta([t], y) \wedge \neg\mathfrak{S}x(\Delta([t], x) \wedge y \neq x))$	Generalization
$\Delta([t], \iota y\Delta([t], y)$	Definite description principle
$t = \iota y\Delta([t], y))$	**Denotation**
$t = ft$	By 1 and SI

Given the Law of Excluded Middle, the line (⋆) follows from **Denotation**.[19] But the argument fails just as much as the one for indefinite descriptions. Not only does it use SI at the last line, but it uses

[19] The proof is as follows. By the LEM, $\neg\Delta([t], x) \vee \Delta([t], x)$. By **Denotation**, $\neg\Delta([t], x) \vee t = x$. So $\neg(\Delta([t], x) \wedge t \neq x)$, and $\mathfrak{A}x\neg(\Delta([t], x) \wedge t \neq x)$ by Universal Generalization. Hence, $\neg\mathfrak{S}y(\Delta([t], x) \wedge t \neq x)$.

Generalization in an illicit form as well, where multiple occurrences of a term are involved.

But, it may be thought, the definite description involved has a unique description, and the term t is obtained by applying the function symbol f to the description. Hence, it has a unique denotation, too. So both these moves are acceptable.

Given the semantics, though, definite descriptions may not have unique denotations. For a start, even if some unique thing satisfies $A(x)$, $\iota x A(x)$ may denote not only this, but some other things too. This might be thought to be an artefact of the denotation conditions given, which ought to be changed. But denotation is not unique for a more fundamental reason. If no unique thing satisfies $A(x)$, then $\iota x A(x)$ still denotes a possible multiplicity of objects. It might be replied that the argument establishes that $\mathfrak{S}y(\Delta([t], y) \wedge \neg\mathfrak{S}x(\Delta([t], x) \wedge t \neq x))$, and hence that we are not in this case. But this is inferred by Generalization on t, so the conclusion is guaranteed only if t has a single denotation. t is obtained by applying a function symbol to the description in question, and so is guaranteed a unique denotation only if the description is. The argument that denotation is unique is therefore circular and question-begging.

Perhaps, it might be thought, we can modify the description in question in such a way that the non-uniqueness case never arises. Thus, let us write $\mathfrak{S}!zA(z)$ for $\mathfrak{S}z(A(z) \wedge \neg\mathfrak{S}x(A(x) \wedge x \neq z))$ and define $\iota^\star$ as follows:

$$\iota^\star x A(x) \text{ is } \varepsilon x(\,(\mathfrak{S}!zA(z) \wedge A(x)) \vee (\neg\mathfrak{S}!zA(z) \wedge x = 0))$$

Since '0' has a unique denotation, $\iota^\star x A(x)$ has a unique denotation in either case.

But now the argument breaks down in another place, as it must in virtue of the non-triviality argument.[20] The inference $\mathfrak{S}!xA(x) \vdash A(\iota^\star x A(x))$ is invalid (even when not substituting into the scope of an ε-term).

To see this, consider its instance $\mathfrak{S}!xPx \vdash P\iota^\star xPx$. Choose an interpretation whose domain is the natural numbers, $\mathcal{N}$; and where $\delta^+(P) = \{1\}$ and $\delta^-(P) = \mathcal{N}$. Since $\Vdash^+_{s(z/1)} \mathfrak{A}x\neg Px$, $\Vdash^+_{s(z/1)} \mathfrak{A}x(\neg Px \vee x = z)$; so $\Vdash^+_{s(z/1)} \neg\mathfrak{S}x(Px \wedge x \neq z)$ and $\Vdash^+_{s(z/1)} Pz \wedge \neg\mathfrak{S}x(Px \wedge x \neq z)$. Hence, $\Vdash^+_s \mathfrak{S}z(Pz \wedge \neg\mathfrak{S}x(Px \wedge x \neq z))$, i.e., $\Vdash^+_s \mathfrak{S}!zPz$. But since $\Vdash^+_s \mathfrak{A}z\neg Pz$, $\Vdash^+_s \mathfrak{A}z(\neg Pz \vee \mathfrak{S}x(Px \wedge x \neq z))$, that is, $\Vdash^+_s \neg\mathfrak{S}z(Pz \wedge \neg\mathfrak{S}x(Px \wedge x \neq z))$,

[20] The situation is similar to that concerning extended paradoxes considered in Priest (1997*b*).

i.e. $\Vdash_s^+ \neg\mathfrak{S}!zPz$. Hence, $\delta(\iota^\star xPx)$ may be $\{0\}$, in which case we do not have $\Vdash_s^+ P\iota^\star xPx$.

As is clear, the problem is that, though we have:

$$(\mathfrak{S}!xPx \wedge P\iota^\star xPx) \vee (\neg\mathfrak{S}!xPx \wedge \iota^\star xPx = 0)$$

$\mathfrak{S}!xPx$ may be both true and false, in which case the second disjunct may hold. Proof-theoretically, the disjunctive syllogism, which is what it would take to get us to the first disjunct, is invalid in a paraconsistent context.

This raises the possibility of extending the language to provide what would be needed to express unique satisfaction consistently. Thus, we might hope to extend the language with a new quantifier, $\mathfrak{S}!!$, which does this. The situation is then very similar to that concerning Boolean negation. Boolean negation, \$, is an operator which, supposedly, behaves as classical negation does. In fact, $\mathfrak{S}!!xA(x)$ can be defined in terms of Boolean negation: $\mathfrak{S}x(\$\$A(x) \wedge \$\mathfrak{S}y(A(y) \wedge \$x = y))$. (The first conjunct consistently expresses the fact that something satisfies $A(x)$; the second conjunct consistently expresses the fact that nothing else does.) So let me discuss Boolean negation.[21]

Why should we suppose that \$ is a coherent notion? This depends on how it is characterized. There are two options. The first is axiomatic. We might simply lay down a set of axioms and/or rules for classical negation and insist that the connective \$ satisfy them. This is not a very satisfactory way to proceed, however. It is open to the objection that \$, so characterized, is a honky connective: it has no determinate sense. We know well that an arbitrary set of postulates may fail to give an operator any determinate sense. Prior and *tonk* have made us all too aware of this possibility.[22] We have a case of the honky tonk blues.

The other possibility is to set up the operator semantically, specifying truth and falsity conditions as:

$\Vdash_s^+ \$A$ iff it is not the case that $\Vdash_s^+ A$

$\Vdash_s^- \$A$ iff $\Vdash_s^+ A$

The question now is how to understand the metalinguistic 'it is not the case that'. It may be the negation of the object language (which is what, for the sake of coherence, one would expect). But in that case, there is no guarantee that \$ will behave classically, since we may well have both

[21] The situation is discussed in Priest (1990), which may be consulted for further details.

[22] See Prior (1960).

$\Vdash_s^+$ \$$A$ and $\Vdash_s^-$ \$$A$. Alternatively, we can insist that it behave classically. But, since we are in the process of giving an argument for the coherence of a notion that satisfies the behaviour or classical negation, this clearly begs the question.

8.9 Conclusion

In this chapter we have looked at what I take to be the hardest objection to noneism, a paradox of denotation that depends crucially on every term having a denotation. This has occasioned a brief excursus into the territory of the paradoxes of self-reference. I gave a logic of multiple denotation, and we saw that the argument can be demonstrably avoided if we take the offending terms to have multiple denotations.

With this, it seems to me, the major final objection to noneism falls away. I am not foolish enough to suppose that there are no others, or that what I have said about the objections that I have discussed is definitive. There is, I am sure, a great deal more to be said about all these issues. But the aim of this book has not been to try to settle the matter of intentionality definitively. That, I am sure, would be a naive aim. I hope that the book has, however, provided the basis of a noneist theory of intentionality, and has taken us, in that sense, towards non-being.

If the theory—or anything like it—is right, then, though actualists have denied it, non-being—what is not—has a determinate and important structure. The structure is, as we have seen, central to understanding many things, and, crucially, intentionality. Intentionality is, as I observed at the beginning of Ch. 1, a fundamental feature of cognitive agents, agents that are. To understand being one has to understand non-being.

8.10 Technical Appendix

In this appendix, I will establish the technical claims referred to in the rest of the chapter. The first couple of proofs establish useful facts about variables.

Lemma 16 *Fix any interpretation. Let t and A be any term and formula. Then if s_1 and s_2 are any evaluations of the variables that agree on the variables free* *in t and A:*

1. $\delta_{s_1}(t) = \delta_{s_2}(t)$
2. $\Vdash^{\pm}_{s_1} A \Leftrightarrow \Vdash^{\pm}_{s_2} A$

Proof The proof is by a joint recursion. For 1:
For constants:

$$\delta_{s_1}(c) = \{\delta(c)\} = \delta_{s_2}(c)$$

For variables:

$$\delta_{s_1}(x) = \{s_1(x)\} = \{s_2(x)\} = \delta_{s_2}(x)$$

For function symbols:

$$\begin{aligned} b \in \delta_{s_1}(ft_1 \dots t_n) &\Leftrightarrow \text{for some } a_1 \in \delta_{s_1}(t_1), \dots, a_n \in \delta_{s_1}(t_n), \\ &\qquad b = \delta(f)(a_1, \dots, a_n) \\ &\Leftrightarrow \text{for some } a_1 \in \delta_{s_2}(t_1), \dots, a_n \in \delta_{s_2}(t_n), \\ &\qquad b = \delta(f)(a_1, \dots, a_n) && \text{by IH} \\ &\Leftrightarrow b \in \delta_{s_2}(ft_1 \dots t_n) \end{aligned}$$

Let τ be the description $\varepsilon y A(y)$. Then:

$$\begin{aligned} b \in \delta_{s_1}(\tau) &\Leftrightarrow b \in \varphi_{\overline{\tau}}\{d\colon \Vdash^{+}_{s_1(x/d)} A(x)\} \cup \varphi_{\overline{\tau}}(D) \\ &\Leftrightarrow b \in \varphi_{\overline{\tau}}\{d\colon \Vdash^{+}_{s_2(x/d)} A(x)\} \cup \varphi_{\overline{\tau}}(D) \quad \text{by 2} \end{aligned}$$

For 2: For atomic formulas:

$$\begin{aligned} \Vdash^{\pm}_{s_1} Pt_1 \dots t_n &\Leftrightarrow \text{for some } x_1 \in \delta_{s_1}(t_1), \dots, x_n \in \delta_{s_1}(t_n), \\ &\qquad \langle x_1, \dots, x_n \rangle \in \delta^{\pm}(P) \\ &\Leftrightarrow \text{for some } x_1 \in \delta_{s_2}(t_1), \dots, x_n \in \delta_{s_2}(t_n), \\ &\qquad \langle x_1, \dots, x_n \rangle \in \delta^{\pm}(P) && \text{by 1} \\ &\Leftrightarrow \Vdash^{\pm}_{s_2} Pt_1 \dots t_n \end{aligned}$$

(If P is a propositional parameter, the condition is just $\langle\rangle \in \delta^{\pm}(P, w)$, and so is independent of s altogether.)

The cases for the connectives and quantifiers are as in Lemma 1 of 1.9.

■

The next lemma shows that a sentence containing a term is true if some denotation of the term satisfies the corresponding open sentence.

Specifically:

Lemma 17 *Fix any interpretation. Let $t'(x)$ and $A(x)$ be any term and formula in which x is not in the scope of an ε. Let t be any term that can be freely substituted for x in these. Let s be any evaluation of the free variables, then if $d \in \delta_s(t)$:*

1. $\delta_{s(x/d)}(t'(x)) \subseteq \delta_s(t'(t))$
2. $\Vdash^{\pm}_{s(x/d)} A(x) \Rightarrow \Vdash^{\pm}_{s} A(t)$

Proof The proof is by joint recursion. For 1, since x is not within the scope of an ε, $t'(x)$ is generated from variables, constants, function symbols, and ε-terms in which x is not free. If t' is a constant or a variable other than x then $t'(x)$ and $t'(t)$ are identical, and x is not free in either. Hence, the result follows from Lemma 16. The same applies to ε-terms in which x is not free. If t' is x, we have:

$$\begin{aligned} b \in \delta_{s(x/d)}(x) &\Rightarrow b \in \{d\} \\ &\Rightarrow b = d \\ &\Rightarrow b \in \delta_s(t) \end{aligned}$$

Let t' be $ft_1(x)\ldots t_n(x)$. Then $b \in \delta_{s(x/d)}(ft_1(x)\ldots t_n(x))$:

$$\begin{aligned} &\Rightarrow \text{for some } a_1 \in \delta_{s(x/d)}(t_1(x)), \ldots, \\ &\qquad a_n \in \delta_{s(x/d)}(t_n(x)), b = \delta(f)(a_1, \ldots, a_n) \\ &\Rightarrow \text{for some } a_1 \in \delta_s(t_1(t)), \ldots, \\ &\qquad a_n \in \delta_s(t_n(t)), b = \delta(f)(a_1, \ldots, a_n) \qquad \text{by IH} \\ &\Rightarrow b \in \delta_s(ft_1(t)\ldots t_n(t)) \end{aligned}$$

For 2: Consider, first, atomic sentences, of the form $Pt_1(x)\ldots t_n(x)$.[23] In each of the t_is x is not in the scope of an ε. Hence, $\Vdash^{\pm}_{s(x/d)} Pt_1(x)\ldots t_n(x)$:

$$\begin{aligned} &\Rightarrow \text{for some } x_1 \in \delta_{s(x/d)}(t_1(x)), \ldots, x_n \in \delta_{s(x/d)}(t_n(x)), \\ &\qquad \langle x_1, \ldots, x_n \rangle \in \delta^{\pm}(P) \\ &\Rightarrow \text{for some } x_1 \in \delta_s(t_1(t)), \ldots, x_n \in \delta_s(t_n(t)), \\ &\qquad \langle x_1, \ldots, x_n \rangle \in \delta^{\pm}(P) \qquad \text{by 1} \\ &\Rightarrow \Vdash^{\pm}_{s} Pt_1(t)\ldots t_n(t) \end{aligned}$$

The cases for the connectives are straightforward.

[23] When P is a propositional parameter, substitution is vacuous, and the result follows from Lemma 16.

Finally, for quantifiers: If $A(x)$ is $\mathfrak{S}yB(x)$, the argument is as follows. The case for the universal quantifier is similar. If x and y are the same, then $A(x)$ and $A(t)$ are the same, and the result follows from Lemma 16. If x and y are distinct:

$$\begin{aligned} \Vdash^{\pm}_{s(x/d)} \mathfrak{S}yB(x) &\Rightarrow \text{for some/all } b \in D, \Vdash^{\pm}_{s(x/d,y/b)} B(x) \\ &\Rightarrow \text{for some/all } b \in D,\ \Vdash^{\pm}_{s(y/b)} B(t) \qquad (\star) \\ &\Rightarrow \Vdash^{\pm}_{s} \mathfrak{S}yB(t) \end{aligned}$$

For ($\star$): Since no variable in t is bound on substitution, t cannot contain y free. By Lemma 16, $\delta_{s(y/b)}(t) = \delta_s(t)$, so we can apply the Induction Hypothesis where s is $s(y/b)$. ■

Corollary 18 *Provided that x is not within the scope of an ε in $A(x)$, and $\varepsilon xA(x)$ and t are free when substituted for x:*

1. $\mathfrak{A}xA(x) \models A(t)$
2. $\mathfrak{S}xA(x) \models A(\varepsilon xA(x))$

Proof For 1: Suppose that $\Vdash^{+}_{s} \mathfrak{A}xA(x)$. Then for all $a \in D, \Vdash^{+}_{s(x/a)} A(x)$. But now choose $a \in \delta_s(t)$. Then $\Vdash^{+}_{s} A(t)$, by Lemma 17.

For 2: Suppose that $\Vdash^{+}_{s} \mathfrak{S}xA(x)$. Then $\{d : \Vdash^{+}_{s(x/d)} A(x)\} \neq \phi$. Choose any $a \in \varphi_{\overline{\varepsilon xA(x)}}\{d : \Vdash^{+}_{s(x/d)} A(x)\}$. Then $a \in \delta_s(\varepsilon xA(x))$ and $\Vdash^{+}_{s(x/a)} A(x)$. By Lemma 17, $\Vdash^{+}_{s} A(\varepsilon xA(x))$. ■

We next show that these inferences have only restricted validity.

Lemma 19 *The inferences of the previous Corollary may be invalid if substitution occurs within the scope of an ε.*

Proof To see that the first inference may fail consider:

$$\mathfrak{A}xP\varepsilon yRxy \vdash P\varepsilon yRty$$

Choose an interpretation where:

$$\begin{aligned} D &= \{a, b, c\} \\ \delta^{+}(P) &= \{a, c\} \\ \delta^{+}(R) &= \{\langle a, a\rangle, \langle a, b\rangle, \langle b, a\rangle, \langle b, c\rangle, \langle c, b\rangle, \langle c, c\rangle\} \end{aligned}$$

To visualize what follows, it may help to keep the following picture of $\delta^+(R)$ in mind:

$$\overset{\curvearrowright}{a} \quad \leftrightarrow \quad b \quad \leftrightarrow \quad \overset{\curvearrowleft}{c}$$

t is a closed term such that:[24]

$$\delta_s(t) = \{a, b, c\}$$

For any t', $\varphi_{t'}$ is independent of t', so we may drop the subscript, and:

$$\begin{aligned} \varphi\{a, b\} &= \{a, b\} \\ \varphi\{b, c\} &= \{b, c\} \\ \varphi\{a, c\} &= \{a, c\} \\ \varphi\{a, b, c\} &= \{b\} \end{aligned}$$

Then:

$$\begin{aligned} \delta_{s(x/a)}(\varepsilon yRxy) &= \varphi\{a, b\} = \{a, b\} \\ \delta_{s(x/b)}(\varepsilon yRxy) &= \varphi\{a, c\} = \{a, c\} \\ \delta_{s(x/c)}(\varepsilon yRxy) &= \varphi\{b, c\} = \{b, c\} \end{aligned}$$

So if $d \in D$, for some $e \in \delta_{s(x/d)}(\varepsilon yRxy)$, $e \in \delta^+(P)$. That is, $\Vdash^+_s \mathfrak{A}xP\varepsilon yRxy$.

But $\delta_s(\varepsilon yRty) = \varphi\{e : \Vdash^+_{s(x/e)} Rty\} = \varphi\{e :$ for some $d \in \delta_s(t)$, $\langle d, e\rangle \in \delta^+(R)\} = \varphi\{a, b, c\} = \{b\}$. So there is no $d \in \delta_s(\varepsilon yRty)$ such that $d \in \delta^+(P)$. That is, $\nVdash^+_s P\varepsilon yRty$.

To see that the second inference may fail, consider:

$$\mathfrak{S}xP\varepsilon yRxy \vdash P\varepsilon yR\tau y$$

where τ is $\varepsilon xP\varepsilon yRxy$.

Choose the same interpretation as before. Then: $\delta_{s(x/a)}(\varepsilon yRxy) = \{a, b\}$. So there is some $d \in \delta_{s(x/a)}(\varepsilon yRxy)$ such that $d \in \delta^+(P)$. That is, for some $e \in D$ (namely, a), $\Vdash^+_{s(x/e)} P\varepsilon yRxy$; so $\Vdash^+_s \mathfrak{S}xP\varepsilon yRxy$.

[24] There is such a term, as τ, in the next part of the proof, demonstrates.

But as we have already seen:

$$\delta_{s(x/a)}(\varepsilon yRxy) = \{a, b\}$$
$$\delta_{s(x/b)}(\varepsilon yRxy) = \{a, c\}$$
$$\delta_{s(x/c)}(\varepsilon yRxy) = \{b, c\}$$

Thus, for every $d \in D$, some $e \in \delta_{s(x/d)}(\varepsilon yRxy)$ is such that $e \in \delta^+(P)$. So $\{d\colon \Vdash^+_{s(x/d)} P\varepsilon yRxy\} = \{a, b, c\}$, and $\delta_s(\tau) = \varphi\{a, b, c\} = \{b\}$. So there is no $d \in \delta_s(\tau)$ such that $d \in \delta^+(P)$. That is, $\nVdash^+_s P\varepsilon yR\tau y$. ■

However, the inferences can be made universally valid if the selection function is appropriately constrained.

Lemma 20 *If, for all X and t, $\varphi_t(X) = X$. Then Lemma 17 holds without the restriction concerning substitution within the scope of an ε-term.*

Proof The proof is the same as that of Lemma 17, except that we need to add a clause in the proof of part 1 of the lemma covering the case where the term in which substitution occurs is an ε-term. This goes as follows.

Let t' be the term $\varepsilon yA(x)$. If x and y are the same, then $t'(x)$ and $t'(t)$ are the same, and x is not free in either; the result follows from Lemma 16. So suppose that x and y are distinct.

$$\begin{aligned} b \in \delta_{s(x/d)}(\varepsilon yA(x)) &\Rightarrow b \in \{a\colon \Vdash^+_{s(x/d,y/a)} A(x)\} \cup D \\ &\Rightarrow b \in \{a\colon \Vdash^+_{s(y/a)} A(t)\} \cup D \qquad (\star) \\ &\Rightarrow b \in \delta_s(\varepsilon yA(t)) \end{aligned}$$

Note that the step marked $(\star)$ is determined twice over. It holds simply because of D. But since no variable in t is bound on substitution, t cannot contain y free. By Lemma 16, $\delta_{s(y/a)}(t) = \delta_s(t)$, so we can apply the Induction Hypothesis where s is $s(y/a)$. (So it would make no difference if D were replaced by any other set.) ■

Corollary 21 *If, for all X and t, $\varphi_t(X) = X$, the inferences of Lemma 18 are valid even for substitution within the scope of an ε.*

Proof The proof is the same as that of Corollary 18, but employing Lemma 20 instead of Lemma 17. ■

We saw that particular generalization is not unrestrictedly valid, even when not substituting into the scope of an ε-term. We next show a valid special case.

Lemma 22 *If the only free occurrence of x in $Pt_1 \ldots x \ldots t_n$ is as shown, then $Pt_1 \ldots t \ldots t_n \models \mathfrak{S}xPt_1 \ldots x \ldots t_n$.*

Proof Assume that $\Vdash_s^+ Pt_1 \ldots t \ldots t_n$:

$$
\begin{aligned}
&\Rightarrow \text{for some } a_1 \in \delta_s(t_1), \ldots, a \in \delta_s(t), \ldots,\\
&\qquad a_n \in \delta_s(t_n), \langle a_1, \ldots, a, \ldots, a_n\rangle \in \delta^+(P)\\
&\Rightarrow \text{for some } a_1 \in \delta_s(t_1), \ldots, a \in \delta_{s(x/a)}(x) \ldots,\\
&\qquad a_n \in \delta_s(t_n), \langle a_1, \ldots, a, \ldots, a_n\rangle \in \delta^+(P) \qquad (\star)\\
&\Rightarrow \text{for some } a_1 \in \delta_{s(x/a)}(t_1), \ldots, a \in \delta_{s(x/a)}(t), \ldots,\\
&\qquad a_n \in \delta_{s(x/a)}(t_n), \langle a_1, \ldots, a, \ldots, a_n\rangle \in \delta^+(P)\\
&\qquad\qquad \text{Lemma 16}\\
&\Rightarrow \text{for some } b \in D, \text{ for some } a_1 \in \delta_{s(x/b)}(t_1), \ldots,\\
&\qquad a \in \delta_{s(x/b)}(t), \ldots, a_n \in \delta_{s(x/b)}(t_n), \langle a_1, \ldots,\\
&\qquad a, \ldots, a_n\rangle \in \delta^+(P)\\
&\Rightarrow \text{for some } b \in D,\ \Vdash_{s(x/b)}^+ Pt_1 \ldots x \ldots t_n\\
&\Rightarrow \Vdash_s^+ \mathfrak{S}xPt_1 \ldots x \ldots t_n
\end{aligned}
$$

For $(\star)$, note that $\delta_{s(x/a)}(x) = \{a\}$. ■

We now show that, given the logic of multiple denotation, the principle **Denotation** does not give triviality, even in the context of arithmetic, with all its self-referential powers. Fix the language to be that of first-order arithmetic, plus the description operator and the denotation and existence predicates.

Henceforth we will deal only with interpretations of a restricted kind. First, the domain of the interpretation is the natural numbers, $\mathcal{N}$, and all the arithmetic machinery has its usual meaning. Thus, the constant **0** denotes 0, $+$ denotes addition, etc., $=$ has its usual (classical) interpretation, and the extension of the existence predicate is fixed in some way. Finally, for all τ and X, $\varphi_\tau(X) = X$.[25] The only things that can vary in an interpretation, therefore, are the extension and co-extension of Δ.

[25] Note that making $\varphi_\tau(X)$ as large as possible spreads contradictions as far as possible. This *increases* the significance of the non-triviality proof.

Let $\mathfrak{I}_1$ and $\mathfrak{I}_2$ be interpretations. Let the extension/co-extension of the predicate Δ in $\mathfrak{I}_1$ and $\mathfrak{I}_2$ be $\Delta_1^{\pm}$ and $\Delta_2^{\pm}$, respectively. Define:

$$\mathfrak{I}_1 \preceq \mathfrak{I}_2 \text{ iff } \Delta_1^{+} \subseteq \Delta_2^{+} \text{ and } \Delta_1^{-} \subseteq \Delta_2^{-}$$

We now establish a crucial monotonicity result. Let the denotation functions and truth/falsity relations of $\mathfrak{I}_1$ and $\mathfrak{I}_2$ be δ^1 and δ^2, and $\Vdash^{1,\pm}$ and $\Vdash^{2,\pm}$, respectively. Then:

Lemma 23 (Monotonicity) *If $\mathfrak{I}_1 \preceq \mathfrak{I}_2$ then for any t, A, and s:*

1. $\delta_s^1(t) \subseteq \delta_s^2(t)$
2. $\Vdash_s^{1,\pm} A \Rightarrow \Vdash_s^{2,\pm} A$

Proof The proof is by a joint recursion. For 1, the cases where t is a constant or variable are trivial. If t is $ft_1 \ldots t_n$ then:

$$\begin{aligned} a \in \delta_s^1(ft_1 \ldots t_n) &\Rightarrow \text{for some } a_1 \in \delta_s^1(t_1), \ldots, a_n \in \delta_s^1(t_n), \\ &\qquad a = \delta(f)(a_1, \ldots, a_n) \\ &\Rightarrow \text{for some } a_1 \in \delta_s^2(t_1), \ldots, a_n \in \delta_s^2(t_n), \\ &\qquad a = \delta(f)(a_1, \ldots, a_n) \qquad \text{by IH} \\ &\Rightarrow a \in \delta_s^2(ft_1 \ldots t_n) \end{aligned}$$

Let t be $\varepsilon x B(x)$. Then:

$$\begin{aligned} a \in \delta_s^1(\varepsilon x B(x)) &\Rightarrow a \in \{d : \Vdash_{s(x/d)}^{1,+} B(x)\} \cup D \\ &\Rightarrow a \in \{d : \Vdash_{s(x/d)}^{2,+} B(x)\} \cup D \quad (\star) \\ &\Rightarrow a \in \delta_s^2(\varepsilon x B(x)) \end{aligned}$$

Again, the step $(\star)$ is overdetermined by both D and 2 of the IH.

For 2, the cases for the connectives and quantifiers are straightforward. For atomic formulas, $=$ and E do not change their denotations, so the only case to worry about is Δ. For this, $\Vdash_s^{1,\pm} \Delta(t_1, t_2)$:

$$\begin{aligned} &\Rightarrow \text{for some } x_1 \in \delta_s^1(t_1), x_2 \in \delta_s^1(t_2), \langle x_1, x_2 \rangle \in \Delta_1^{\pm} \\ &\Rightarrow \text{for some } x_1 \in \delta_s^2(t_1), x_2 \in \delta_s^2(t_2), \langle x_1, x_2 \rangle \in \Delta_1^{\pm} \quad \text{by 1} \\ &\Rightarrow \text{for some } x_1 \in \delta_s^2(t_1), x_2 \in \delta_s^2(t_2), \langle x_1, x_2 \rangle \in \Delta_2^{\pm} \quad \text{since } \mathfrak{I}_1 \preceq \mathfrak{I}_2 \\ &\Rightarrow \Vdash_s^{2,\pm} \Delta(t_1, t_2) \end{aligned}$$

■

We now construct a fixed-point interpretation that is a model of **Denotation**. We define a sequence of interpretations, $\mathfrak{I}_\alpha$, for every ordinal α. Let the denotation function and truth/falsity relations of $\mathfrak{I}_\alpha$ be δ^α and $\Vdash^{\alpha,\pm}$, respectively; and write $\delta^{\alpha,\pm}(\Delta)$ as $\Delta^\pm_\alpha$. We suppose that we have some arithmetic coding of terms and formulas. Let the code of t be $\#t$, and $[t]$ be the numeral of this. If a is not the code of a closed term, then for any b and α, $\langle a, b\rangle$ is in Δ^-_α but not in Δ^+_α. Hence, we need concern ourselves only with matters when a is the code of a closed term, t. The definition is by recursion.

- For all b, $\langle \#t, b\rangle \in \Delta^+_0$ and $\langle \#t, b\rangle \in \Delta^-_0$.
- For successor ordinals:
  $\langle \#t, b\rangle \in \Delta^\pm_{\alpha+1}$ iff $\Vdash^{\alpha,\pm}_{s(x/b)} t = x$
- If λ is a limit ordinal:
  $\Delta^\pm_\lambda = \bigcap_{\alpha<\lambda} \Delta^\pm_\alpha$

Lemma 24 *For all α, β, if $\alpha \leq \beta$, $\mathfrak{I}_\alpha \succeq \mathfrak{I}_\beta$.*

Proof The proof is by recursion on α. Since the behaviour of Δ does not change with respect to pairs whose first members are not the codes of closed terms, we need concern ourselves only with ones that are. If α is 0 the result is trivial. Let α be a limit ordinal, λ, and suppose that $\langle \#t, b\rangle \in \Delta^\pm_\beta$. By induction hypothesis, for all $\gamma < \lambda$, $\langle \#t, b\rangle \in \Delta^\pm_\gamma$. Hence, $\langle \#t, b\rangle \in \Delta^\pm_\lambda$. Now suppose that α is a successor, $\gamma + 1$. If β is α, or a limit ordinal, the result is trivial. So suppose that $\beta = \eta + 1$ and $\langle \#t, b\rangle \in \Delta^\pm_\beta$. By definition, $\Vdash^{\eta,\pm}_{s(x/b)} t = x$. By induction hypothesis, $\mathfrak{I}_\gamma \succeq \mathfrak{I}_\eta$, so by Lemma 23, $\Vdash^{\gamma,\pm}_{s(x/b)} t = x$, so $\langle \#t, b\rangle \in \Delta^\pm_\alpha$. ∎

Note that the Lemma shows that for each α, $\mathfrak{I}_\alpha$ is indeed an *LP* interpretation (that is $\Delta^+_\alpha \cup \Delta^-_\alpha = \mathcal{N}^2$). This is obviously true for 0. If it is true for α, it is true for $\alpha + 1$, since the LEM holds in *LP*. And if λ is a limit ordinal, and it is true for all $\alpha < \lambda$, it is true for λ. For suppose that $x \notin \Delta^+_\lambda \cup \Delta^-_\lambda$; then for some $\alpha, \beta < \lambda$, $x \notin \Delta^+_\alpha$ and $x \notin \Delta^-_\beta$. If γ is the greater of α and β, then, by the Lemma, $x \notin \Delta^+_\gamma \cup \Delta^-_\gamma$. The result follows by transfinite induction.

By standard set-theoretic considerations, the extension and co-extension of Δ cannot keep decreasing as we ascend the ordinals. There must therefore be an ordinal, θ, such that $\mathfrak{I}_\theta = \mathfrak{I}_{\theta+1}$.

Lemma 25 *$\mathfrak{I}_\theta$ (i.e., $\mathfrak{I}_{\theta+1}$) verifies* ***Denotation.***

Proof Note that, for any closed t, since $[t]$ is a numeral, it has a unique denotation, namely $\#t$. Then $\Vdash_s^{\theta+1,\pm} \Delta([t], x)$:

$$\begin{aligned}
&\Leftrightarrow \text{for some } a \in \delta_s^{\theta+1}([t]), b \in \delta_s^{\theta+1}(x), \langle a, b\rangle \in \Delta^{\pm}_{\theta+1}\\
&\Leftrightarrow \text{for some } b \in \delta_s^{\theta+1}(x), \langle \#t, b\rangle \in \Delta^{\pm}_{\theta+1}\\
&\Leftrightarrow \text{for some } b \in \delta_s^{\theta}(x), \Vdash_{s(x/b)}^{\theta,\pm} t = x\\
&\Leftrightarrow \text{for some } b \in \delta_s^{\theta}(x), a \in \delta_{s(x/b)}^{\theta}(t), c \in \delta_{s(x/b)}^{\theta}(x), a = c\\
&\Leftrightarrow \text{for some } b \in \delta_s^{\theta}(x), a \in \delta_{s(x/b)}^{\theta}(t), c \in \{b\}, a = c\\
&\Leftrightarrow \text{for some } b \in \delta_s^{\theta}(x), a \in \delta_{s(x/b)}^{\theta}(t), a = b\\
&\Leftrightarrow \text{for some } b \in \delta_s^{\theta}(x), a \in \delta_s^{\theta}(t), a = b \qquad (\star)\\
&\Leftrightarrow \Vdash_s^{\theta,\pm} t = x\\
&\Leftrightarrow \Vdash_s^{\theta+1,\pm} t = x
\end{aligned}$$

For $(\star)$, t is closed, so this follows by Lemma 16. ∎

Theorem 26 *There is an inconsistent non-trivial theory in the logic of multiple denotation containing arithmetic, and satisfying* ***Denotation*** *(and* ***Des*** *and Universal Instantiation in the unrestricted form of Corollary 21).*

Proof Consider the set of sentences that hold in $\mathfrak{I}_\theta$. This contains **Denotation** by the previous Lemma. It is non-trivial, since every purely arithmetic sentence takes its classical value. To see that it is inconsistent, consider the term $\varepsilon x\, x = x$. Call this t. Everything satisfies $x = x$. Hence, t denotes 0, 1 (and everything else). Since 0 satisfies $x = \mathbf{0}$, and 1 satisfies $\neg x = \mathbf{0}$, $t = \mathbf{0}$ and $\neg t = \mathbf{0}$ hold in the theory, by Lemma 17. Finally, because of the definition of φ in the interpretation, **Des** and Universal Instantiation hold generally. ∎

Bibliography

Anderson, C. A., and Owens, J. (eds.) (1990), *Propositional Attitudes: the Role of Content in Logic, Language and Mind*, CSLI Lecture Notes 20 (Stanford, Calif.: CSLI).

Ashworth, E. J. (1977), 'Chimeras and Imaginary Objects: A Study in Post-Medieval Theory of Signification', *Vivarium* 15: 57–79.

Balaguer, M. (1995), 'A Platonist Epistemology', *Synthese* 103: 303–25.

Beall, J. C. (1999), 'From Full Blooded Platonism to Really Full Blooded Platonism', *Philosophia Mathematica* 7: 322–5.

—— Brady, R., Hazen, A., Priest, G., and Restall, G. (in preparation), 'Restricted Quantification'.

Boër, S. E., and Lycan, W. G. (1986), *Knowing Who* (Cambridge, Mass.: MIT Press).

Boh, I. (1993), *Epistemic Logic in the Later Middle Ages* (London: Routledge).

Bos, E. P. (1983), *Marsilius of Inghen: Treatises on the Properties of Terms* (Dordrecht: Reidel).

Brentano, F. (1874), *Psychologie vom Empirische Standpunkt*, tr. A. C. Rancurello, D. B. Terrell, and L. L. McAlister, as *Psychology from an Empirical Standpoint* (London: Routledge & Kegan Paul, 1973).

Bressan, A. (1972), *A General Interpreted Modal Calculus* (New Haven, Conn.: Yale University Press).

Burgess, J. P., and Rosen, G. (1997), *A Subject with No Object* (Oxford: Clarendon Press).

Buridan, J. (1976), *Tractatus de Consequentiis*, ed. H. Hubien (Louvain: Publications Universitaires, and Paris: S.A. Vander-Oyez).

—— (2001), *Summulae de Dialectica*, tr. G. Klima (New Haven, Conn.: Yale University Press).

Chisholm, R. (ed.) (1960), *Realism and the Background to Phenomenology* (London: Allen & Unwin).

—— (1967), 'Identity Through Possible Worlds: Some Questions', *Noûs* 1: 1–8; repr. in M. J. Loux (ed.), *The Possible and the Actual* (Ithaca, NY: Cornell University Press), ch. 3.

Crane, T. (1995), *The Mechanical Mind: a Philosophical Introduction to Minds, Machines and Mental Representation* (London: Penguin Books).

—— (1998), 'Intentionality', in E. Craig (ed.), *The Routledge Encyclopedia of Philosophy* (London: Routledge), v. 816–21.

Currie, G. (1990), *The Nature of Fiction* (Cambridge: Cambridge University Press).

De Rijk, L. M. (1982), 'The Origins of the Theory of the Properties of Terms', in N. Kretzmann *et al*. (eds.), *The Cambridge History of Later Medieval Philosophy* (Cambridge: Cambridge University Press), 161–73.

Devitt, M. (1983), 'Realism and the Renegade Putnam', *Noûs* 17: 291–30.

—— and Sterelny, K. (1987), *Language and Reality: An Introduction to the Philosophy of Language* (Oxford: Basil Blackwell).

Ebbesen, S. (1986), 'The Chimera's Diary', in S. Knuuttila and J. Hintikka (eds.), *The Logic of Being* (Dordrecht: Reidel), 115–43.

Edelberg, W. (1986), 'A New Puzzle about Intentional Identity', *Journal of Philosophical Logic* 15: 1–25.

Field, H. (1998), 'Which Undecidable Sentences have Truth Values?', in H. G. Dales and G. Olivieri (eds.), *Truth in Mathematics* (Oxford: Clarendon Press), ch. 16.

Fine, K. (1982), 'The Problem of Non-Existents I—Internalism', *Topoi* 1: 97–140.

Fitting, M., and Mendelsohn, R. (1998), *First Order Modal Logic* (Dordrecht: Kluwer Academic Publishers).

Frege, G. (1952), 'On Sense and Reference', in P. Geach and M. Black (trs.), *Translations from the Writings of Gottlob Frege* (Oxford: Basil Blackwell).

Geach, P. (1967), 'Intentional Identity', *Journal of Philosophy* 64: 627–32; repr. as sect. 4.4 of *Logic Matters* (Oxford: Basil Blackwell, 1972).

Griffin, N. (1998), 'Problems in Item Theory', a paper read at the 1998 meeting of the Australasian Association for Logic.

Gupta, A. (1980), *The Logic of Common Nouns* (New Haven, Conn.: Yale University Press).

Hicks, R. D. (tr.) (1925), *Diogenes Laertius: Lives of Eminent Philosophers* (Cambridge, Mass.: Harvard University Press).

Hilbert, D., and Bernays, P. (1939), *Grundlagen der Mathematik* (Berlin: Springer Verlag), ii.

Hintikka, J. (1962), *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Ithaca, NY: Cornell University Press).

—— (1969), 'Semantics for Propositional Attitudes', in J. W. Davis *et al*. (eds.), *Philosophical Logic* (Dordrecht: Reidel), 21–95; repr. in Linsky (1971), ch. 10.

Horcutt, M. O. (1972), 'Is Epistemic Logic Possible?', *Notre Dame Journal of Formal Logic* 13: 433–53.

Howell, R. (1998), 'Fiction, Semantics of', in E. Craig (ed.), *The Routledge Encyclopedia of Philosophy* (London: Routledge), iii. 659–63.

Hughes, G. E., and Cresswell, M. (1968), *Introduction to Modal Logic* (London: Methuen).

Klima, G. (1991), '*Debeo Tibi Equum*: A Reconstruction of the Theoretical Framework of Buridan's Treatment of the Sophismata', in S. Read (ed.), *Sophisms in Medieval Logic and Grammar: Proceedings of the 9th European Symposium for Medieval Logic and Semantics* (Dordrecht: Kluwer Academic Publishers), 333–47.

KNEALE, W., and KNEALE, M. (1962), *The Development of Logic* (Oxford: Clarendon Press).

KRIPKE, S. (1972), 'Naming and Necessity', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language* (Dordrecht: Reidel), 253–355; repr. as *Naming and Necessity* (Oxford: Basil Blackwell, 1980).

—— (1977), 'Identity and Necessity', in S. P. Schwartz (ed.), *Naming, Necessity and Natural Kinds* (Ithaca, NY: Cornell University Press), ch. 2.

—— (1979), 'A Puzzle about Belief', in A. Margalit (ed.), *Meaning and Use* (Dordrecht: Reidel), 239–83.

LACKEY, D. (ed.) (1973), *Essays in Analysis* (London: Allen & Unwin).

LAKOFF, G. (1970), 'Linguistics and Natural Logic', *Synthese* 22: 151–71.

LEISENRING, A. C. (1969), *Mathematical Logic and Hilbert's ε-Symbol* (London: Macdonald Technical and Scientific).

LEWIS, D. (1968), 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 65: 113–26; repr. in Lewis (1983), ch. 3.

—— (1978) 'Truth in Fiction', *American Philosophical Quarterly* 15: 37–46; repr. in Lewis (1983), ch. 15.

—— (1983), *Philosophical Papers* (Oxford: Oxford University Press), i.

—— (1984), 'Putnam's Paradox', *Australasian Journal of Philosophy* 62: 221–36; repr. in Lewis (1999), ch. 2.

—— (1986), *On the Plurality of Worlds* (Oxford: Blackwell).

—— (1990), 'Noneism or Allism', *Mind* 99: 23–31; repr. in Lewis (1999), ch. 8.

—— (1999), *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press).

LINSKY, L. (1971), *Reference and Modality* (Oxford: Oxford University Press).

MAIERÙ, A. (1972), *Terminologia Logica della Tarda Scolastica* (Rome: Edizioni dell'Ateneo).

MEINONG, A. (1904), '*Gegenstandstheorie*', in A. Meinong (ed.), *Untersuchungen zur Gegenstandstheorie und Psychologie* (Leipzig). Translated into English as 'The Theory of Objects', in Chisholm (1960), ch. 4; the most important part of this is reprinted as 'Kinds of Being', in G. Iseminger (ed.), *Logic and Philosophy; Selected Readings* (New York, NY: Appleton-Century Crofts, 1968), 116–27.

NICHOLS, R. (2002), 'Reid on Fictional Objects and the Way of Ideas', *Philosophical Quarterly* 52: 582–601.

NOLAN, D. (1998), 'An Uneasy Marriage', a talk given at the 1998 meeting of the Australasian Association of Philosophy.

OCKHAM, G. DE (1974), *Ockham's Theory of Terms* (*Part 1 of the* Summa Logicae), tr. M. J. Loux, (Notre Dame, Ind.: University of Notre Dame Press).

—— (1979), *Expositio super Libros Elenchorum*, ed. F. del Punta (St Bonaventure, NY: Franciscan Institute Publications).

PARKS, Z. (1974), 'Semantics for Contingent Identity Systems', *Notre Dame Journal of Formal Logic* 15: 333–4.

—— (1976), 'Investigations into Quantified Modal Logic', *Studia Logica* 35: 109–25.

Parks, Z. and Smith, T. L. (1974), 'The Inadequacy of Hughes and Cresswell's Semantics for Contingent Identity Systems', *Notre Dame Journal of Formal Logic* 15: 331–2.

Parsons, T. (1980), *Nonexistent Objects* (New Haven, Conn.: Yale University Press).

Paul of Venice (1499), *Logica Magna* (Venice: Bonetus Locatellus for Octavianus Scotus).

——(1971), *Logica Magna: Tractatus de Suppositionibus*, ed. and tr. A. Perreiah (St Bonaventure, NY: The Franciscan Institute).

——(1978), *Logica Magna: Secunda Pars, Tractatus de Veritate et Falsitate Propositionis et Tractatus de Significatio Propositionis*, ed. F. del Punta, tr. M. M. Adams (Oxford: Oxford University Press for the British Academy).

——(1981), *Logica Magna: Prima Pars, Tractatus de Scire et Dubitare*, ed. and tr. P. Clarke (Oxford: Oxford University Press for the British Academy).

——(1990), *Logica Magna: Secunda Pars, Capitula de Conditionali et de Rationali*, ed. and tr. G. E. Hughes (Oxford: Oxford University Press for the British Academy).

Peter of Spain (1972), *Tractatus* (called afterwards *Summulae Logicales*), ed. L. M. De Rijk (Assen: van Gorcum).

Priest, G. (1979), 'Indefinite Descriptions', *Logique et Analyse* 22: 5–21.

——(1987), *In Contradiction* (Dordrecht: Martinus Nijhoff); 2nd (extended) edn. (Oxford: Oxford University Press, forthcoming).

——(1990), 'Boolean Negation and All That', *Journal of Philosophical Logic* 19: 201–15; substantially repr. in *Doubt Truth to be a Liar* (Oxford: Oxford University Press, forthcoming), ch. 5.

——(1995*a*), *Beyond the Limits of Thought* (Cambridge: Cambridge University Press); 2nd (extended) edn. (Oxford: Oxford University Press, 2002).

——(1995*b*), 'Multiple Denotation, Ambiguity, and the Strange Case of the Missing Amoeba', *Logique et Analyse* 38: 361–73.

——(1997*a*), 'Sylvan's Box', *Notre Dame Journal of Formal Logic* 38: 573–82.

——(1997*b*), 'On a Paradox of Hilbert and Bernays', *Journal of Philosophical Logic* 26: 45–56.

——(1997*c*), 'Inconsistent Models of Arithmetic, I: Finite Models', *Journal of Philosophical Logic* 26: 223–35.

——(1998), 'Fuzzy Identity and Local Validity', *Monist* 81: 331–42.

——(1999), 'Semantic Closure, Descriptions and Non-Triviality', *Journal of Philosophical Logic* 28: 549–58.

——(2000*a*), 'Objects of Thought', *Australasian Journal of Philosophy* 78: 494–502.

——(2000*b*), 'Inconsistent Models of Arithmetic, II: the General Case', *Journal of Symbolic Logic* 65: 1519–29.

——(2001), *An Introduction to Non-Classical Logic* (Cambridge: Cambridge University Press).

——(2002), 'The Hooded Man', *Journal of Philosophical Logic* 31: 445–67.

——(2003), 'Meinong and the Philosophy of Mathematics', *Philosophia Mathematica* 11: 3–15.

——(Forthcoming *a*), 'The Paradoxes of Denotation', in T. Bolander, V. F. Hendricks, and S. A. Pedersen (eds.), *Self-Reference* (Stanford, Calif.: CSLI Publications).

——(Forthcoming *b*), 'Conditionals: a Debate with Jackson', in I. Ravenscroft (ed.), *Minds, Worlds and Conditionals: Themes from the Philosophy of Frank Jackson* (Oxford: Oxford University Press).

——(Forthcoming *c*), 'Non-Transitive Identity', a paper presented to the third Word Conference on Paraconsistency, Toulouse, 2003.

——and Read, S. (1981), 'Ockham's Rejection of Ampliation', *Mind* 90: 274–9.

————(2004), 'Intentionality—Meinongianism and the Medievals', *Australasian Journal of Philosophy* 82: 416–35.

——Routley, R., and Norman, J. (eds.) (1989), *Paraconsistent Logic: Essays on the Inconsistent* (Munich: Philosophia Verlag).

Prior, A. (1960), 'The Runabout Inference Ticket', *Analysis* 21: 38–9; repr. in P. Strawson (ed.), *Philosophical Logic* (Oxford: Oxford University Press, 1967), 129–31.

Putnam, H. (1980), 'Models and Reality', *Journal of Symbolic Logic* 45: 464–82; repr. in *Realism and Reason: Philosophical Papers* (Cambridge: Cambridge University Press, 1983), iii. ch. 1.

Quine, W. V. O. (1948), 'On What There Is', *Review of Metaphysics* 48, 21–38; repr. as ch. 1 of *From a Logical Point of View*, Cambridge, MA: Harvard University Press.

——(1956), 'Quantifiers and Propositional Attitudes', *Journal of Philosophy* 53: 177–87; repr. in Linsky (1971), ch. 8.

Rantala, V. (1982), 'Impossible World Semantics and Logical Omniscience', *Acta Philosophica Fennica* 35: 106–15.

Read, S. (2001), 'Medieval Theories of Properties of Terms', *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/medieval-terms/, accessed 11 Oct. 2004.

Ross, W. D. (ed. and tr.) (1928), *The Works of Aristotle* (Oxford: Clarendon Press).

Routley, R. (1980), *Exploring Meinong's Jungle and Beyond* (Canberra: RSSS, Australian National University).

——(1982), 'On What There Isn't', *Philosophy and Phenomenological Research* 43: 151–78; also in Routley (1980), ch. 3.

——(2003), 'The Importance of Nonexistent Objects and of Intensionality in Mathematics', *Philosophia Mathematica* 11: 20–52; parts of Routley (1980), chs. 10 and 11.

——Plumwood, V., Meyer, R. K., and Brady, R. T. (1982), *Relevant Logics and their Rivals* (Atascadero, Calif.: Ridgeview), i.

Russell, B. (1903), *Principles of Mathematics* (Cambridge: Cambridge University Press).

——(1905*a*), Review of A. Meinong, *Untersuchungen zur Gegenstandstheorie und Psychologie*, *Mind* 14: 530–8; repr. in Lackey (1973), ch. 2.

Russell, B. (1905*b*), 'On Denoting', *Mind* 14: 479–93; repr. in Lackey (1973), ch. 5.
——(1919), *Introduction to Mathematical Philosophy* (London: Allen & Unwin).
Ryle, G. (1973), 'Intentionality-Theory and the Nature of Thinking', *Revue Internationale de Philosophie* 27: 255–64.
Salmon, N. (1986), *Frege's Puzzle* (Atascadero, Calif.: Ridgeview).
——(1995), 'Being in Two Minds: Belief with Doubt', *Noûs* 29: 1–20.
——and Soames, S. (1988), *Propositions and Attitudes* (Oxford: Oxford University Press).
Scott, T. K. (ed.) (1977), *Sophismata* (Stuttgart-Bad Canstatt: Frommann-Holzboog).
Searle, J. R. (1983), *Intentionality: An Essay on the Philosophy of Mind* (Cambridge: Cambridge University Press).
Sorabji, R. (1991), 'From Aristotle to Brentano: the Development of the Concept of Intentionality', in H. Blumenthal and H. Robinson (eds.), *Oxford Studies in Ancient Philosophy, Suppl. Vol.: Aristotle and the Later Tradition* (Oxford: Oxford University Press), 227–59.
Walton, K. (1978), 'Fearing Fictions', *Journal of Philosophy* 75: 5–27.
——(1990), *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Cambridge, Mass.: Harvard University Press).
Wang, W-F. (2004), 'A Model-Theoretic Argument against Meinongianism and Modal Realism', in E. Leinfellner *et. al.* (eds.), *Experience and Analysis: Proceedings of the 27th International Wittgenstein Symposium* (Kirchberg am Wechsel: Austrian Ludwig Wittgenstein Society), 403–4.
Yagisawa, T. (1988), 'Beyond Possible Worlds', *Philosophical Studies* 53: 175–204.
Yi, B. (1999), 'Is Two a Property?', *Journal of Philosophy* 96: 163–90.
Zalta, E. (1988), *Intensional Logic and the Metaphysics of Intentionality* (Cambridge, Mass.: MIT Press).

Index