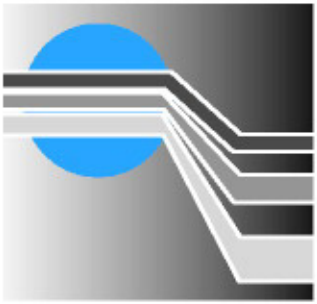
CAMBRIDGE STUDIES IN PHILOSOPHY



MEANING, EXPRESSION AND THOUGHT

WAYNE A. DAVIS

CAMBRIDGE more information - www.cambridge.org/9780521555135

This page intentionally left blank

Meaning, Expression, and Thought

This philosophical treatise on the foundations of semantics is a systematic effort to clarify, deepen, and defend the classical doctrine that words are conventional signs of mental states, principally thoughts and ideas, and that meaning consists in their expression. This expression theory of meaning is developed by carrying out the Gricean program, explaining what it is for words to have meaning in terms of speaker meaning, and what it is for a speaker to mean something in terms of intention. But Grice's own formulations are rejected, and alternatives are developed. The foundations of the expression theory are explored at length, and the author develops the theory of thought as a fundamental cognitive phenomenon distinct from belief and desire and argues for the thesis that thoughts have parts, identifying ideas or concepts with parts of thoughts.

This book will appeal to students and professionals interested in the philosophy of language.

Wayne A. Davis is a professor of philosophy at Georgetown University.

CAMBRIDGE STUDIES IN PHILOSOPHY

General editor ERNEST SOSA (Brown University)

Advisory editors:

JONATHAN DANCY (University of Reading)
JOHN HALDANE (University of St. Andrews)
GILBERT HARMAN (Princeton University)
FRANK JACKSON (Australian National University)
WILLIAM G. LYCAN (University of North Carolina at Chapel Hill)
SYDNEY SHOEMAKER (Cornell University)
JUDITH J. THOMSON (Massachusetts Institute of Technology)

RECENT TITLES

BARRY MAUND *Colours*
MICHAEL DEVITT *Coming to Our Senses*
MICHAEL ZIMMERMAN *The Concept of Moral Obligation*
MICHAEL STOCKER with ELIZABETH HEGEMAN *Valuing Emotions*
SYDNEY SHOEMAKER *The First-Person Perspective and Other Essays*
NORTON NELKIN *Consciousness and the Origins of Thought*
MARK LANCE and JOHN O'LEARY HAWTHORNE *The Grammar of Meaning*
D. M. ARMSTRONG *A World of States of Affairs*
PIERRE JACOB *What Minds Can Do*
ANDRE GALLOIS *The World Without the Mind Within*
FRED FELDMAN *Utilitarianism, Hedonism, and Desert*
LAURENCE BONJOUR *In Defense of Pure Reason*
DAVID LEWIS *Papers in Philosophical Logic*
WAYNE DAVIS *Implicature*
DAVID COCKBURN *Other Times*
DAVID LEWIS *Papers on Metaphysics and Epistemology*
RAYMOND MARTIN *Self-Concern*
ANNETTE BARNES *Seeing Through Self-Deception*
MICHAEL BRATMAN *Faces of Intention*
AMIE THOMASSON *Fiction and Metaphysics*
DAVID LEWIS *Papers on Ethics and Social Philosophy*
FRED DRETSKE *Perception, Knowledge and Belief*
LYNNE RUDDER BAKER *Persons and Bodies*
JOHN GRECO *Putting Skeptics in Their Place*
DERK PEREBOOM *Living Without Free Will*
BRIAN ELLIS *Scientific Essentialism*
RICHARD FOLEY *Intellectual Trust in Oneself and Others*
ALAN GOLDMAN *Practical Rules*

Meaning, Expression, and Thought

WAYNE A. DAVIS
Georgetown University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom

Published in the United States by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521555135

First published in print format 2003

ISBN-13 978-0-511-06779-2 eBook (EBL)
ISBN-10 0-511-06779-8 eBook (EBL)

ISBN-13 978-0-521-55513-5 hardback
ISBN-10 0-521-55513-2 hardback

Dedicated to David K. Lewis, model philosopher,
with deep gratitude.

Contents

Preface

I began work on thought, belief, and desire shortly after I graduated from the University of Michigan in 1973, inspired by Alvin Goldman and his *A Theory of Human Action*, along with Stephen Stich, Arthur Burks, John Perry, and Jaegwon Kim. That work grew into my doctoral dissertation (Princeton University, 1977), directed by David Lewis, Gilbert Harman, and Richard Jeffrey. I remain indebted to these outstanding philosophers not only for key ideas but also for instilling a love of philosophy. The dissertation became a book-length manuscript entitled "Elements of Psychology: Belief, Desire, and Thought." When a chapter on meaning took on the proportions of a book all by itself, I decided to first complete the present volume, *Meaning, Expression, and Thought*. Many of the ideas on thought presented in Part III were first developed in my dissertation and elaborated in "Belief, Desire, and Thought." I use them here to provide the psychological foundations for the theory of meaning developed in the rest of this work. This book was delayed by my recent *Implicature* (1998), which explains why Grice's great "synthetic" project gets so much less attention here than his "analytic" project. I wrote *Meaning, Expression, and Thought*, furthermore, in tandem with my forthcoming *Nondescriptive Meaning and Reference*, which applies the expression theory of meaning to names, indexicals, and other special cases, develops the expression theory of reference in greater depth, and shows how referential semantics can be treated in the expression theory.

While revised and reorganized here, most of the material in Chapters 1–4 has appeared in the following publications: "Expression of Emotion," *American Philosophical Quarterly* 25 (1988): 279–291; "Speaker Meaning," *Linguistics and Philosophy* 15 (1992): 223–253; "Cogitative and Cognitive Speaker Meaning," *Philosophical Studies* 67 (1992): 71–88; and

"Communicating, Telling, and Informing," *Philosophical Inquiry* 21 (1999): 21–43. A summary of Chapter 9, "Conventions," appeared in my *Implicature* (Cambridge University Press, 1998). The rest is new. I am grateful for the editors of these journals for permission to reprint.

In the interest of readability, I adopt a casual approach to the use-mention distinction. I use either italics or quotation marks when using words to refer to themselves or to give meanings. I also use them as corner quotes around variables to form placeholders for quoted or italicized words. And, of course, I use quotation marks for direct quotation and scare quotes, and italics for emphasis. I will ensure that context makes my meaning clear (to the charitable reader). Thus in "*vixen* means 'female fox,'" the meaning of the word *means* dictates that the italicized word to its left refers to a word, and that the quoted words to its right give the meaning of the word referred to.

In footnotes, I use "cf." when citing authors who defend views similar to the one under discussion, and "contrast" when citing authors who reject such views.

My treasured colleagues and graduate students at Georgetown University deserve many thanks, especially Mark Lance, Linda Wetzel, Joseph Rahill, and Matt Burstein. I cannot thank Steven Kuhn enough for going over early drafts of the entire manuscript with a fine-toothed comb, and providing pages and pages of useful and incisive criticism. I am also grateful to Georg Meggle, Mark Siebel, Christian Plunze, Christoph Jäger, Thomas Bartelborth, and Oliver Scholz at the University of Leipzig for the hospitality they showed both me and my ideas. John Hawthorne, Mark Heller, Dan Sperber, Michael Slote, Georges Rey, Adrienne Lehrer, Andrew Milne, Stephen Rieber, and Christoph Doerge, along with Ernie Sosa, Francis J. Pelletier, and numerous reviewers, provided many helpful comments. Russell Hahn did a wonderful job as copyeditor and production editor. I am especially grateful to Robert Audi, William Lycan, Robert and Marilyn Adams, Daniel Robinson, Terry Pinkard, and Tom Beauchamp for their support and friendship over the years. Georgetown University provided the resources that enabled me to do the bulk of my research. Jack Bender gets credit not only for comments and friendship but also for helping to form my psycho-philosophical mind at Michigan. Alan Spiro has provided friendship and support since our Princeton days. Most of all, I am indebted to my wife, Kathy Olesko, for more than twenty-seven years of intellectual stimulation and love.

Terry Moore at Cambridge University Press showed the patience of a saint in waiting for the final manuscript, as did Ernie Sosa, the series

editor. In addition to the vast literature on meaning, expression, and thought that existed before I began work on this book, new literature has been published faster than this human being, at least, could keep up with. It seems that every time I completed a draft, Jerry Fodor published another book. I apologize to the authors as well as to the reader for the many omissions I have found to be inevitable.

As a consequence of their long gestation, many of my ideas have been anticipated in print. Research showed that others were not new at all. While I may have lost the right to claim priority, I have benefited immeasurably from the work of others on "my" ideas. I hope I repay the authors I cite here by taking their ideas further.

Washington, D.C.
July 2001

1

Introduction

One of the most venerable doctrines in the history of philosophy, linguistics, and psychology is the thesis that words are conventional signs of mental states, principally thoughts and ideas, and that meaning consists in their expression. This *expression theory of meaning*, as I call it, is firmly entrenched in our commonsense understanding of the world. But familiarity has bred complacency as well as contempt. Development of the doctrine was limited through the nineteenth century, and the twentieth century brought denunciation of the expression theory from generations of scholars. Behavioristic theories of meaning have now faded from view. But referential theories still dominate the field, despite insurmountable problems. This work is an extended effort to clarify, deepen, and defend the expression theory, thereby systematizing what is known about meaning and expression. The best way to do this, I believe, is to carry out the Gricean program, explaining what it is for words to have meaning in terms of speaker meaning and what it is for a speaker to mean something in terms of intention. To succeed in this project, we must develop the theory of thought as a fundamental mental phenomenon distinct from belief and desire, identifying ideas with parts of thoughts. This work, then, is a philosophical treatise on the foundations of semantics.

§1.1 MEANING AS THE EXPRESSION OF THOUGHT

Like many other central philosophical and scientific ideas, the expression theory was first set out by Plato (427–347 B.C.) and Aristotle (384–322 B.C.).

> Spoken words are the symbols of mental experience and written words are the symbols of spoken words. Just as all men have not the same writing, so all men

have not the same speech sounds, but the mental experiences, which these directly symbolize, are the same for all, as are those things of which our experiences are the images. . . . As there are in the mind thoughts which do not involve truth or falsity, and also those which must be either true or false, so it is in speech. For truth and falsity imply combination and separation. . . . A sentence is a significant portion of speech, some parts of which have an independent meaning, that is to say, as an utterance, though not as the expression of any positive judgement. . . . Every sentence has meaning, not as being the natural means by which a physical faculty is realized, but, as we have said, by convention. Yet every sentence is not a proposition; only such are propositions as have in them either truth or falsity. Thus a prayer is a sentence, but is neither true nor false. (Aristotle, *De Interpretatione*: §§1–4)

Aristotle became "the Philosopher" during the medieval period, and his views were kept alive by Augustine (A.D. 354–430), Boethius (ca. 475–525), Avicenna (ca. 929–1037), and Ockham (ca. 1280–1349).

[A] sign is a thing which, over and above the impression it makes on the senses, causes something else to come into the mind as a consequence of itself. . . . Natural signs are those which, apart from any intention or desire of using them as signs, do yet lead to the knowledge of something else, as, for example, smoke when it indicates fire. . . . Conventional signs, on the other hand, are those which living beings mutually exchange for the purpose of showing, as well as they can, the feelings of their minds, or their perceptions, or their thoughts. Nor is there any reason for giving a sign except the desire of drawing forth and conveying into another's mind what the giver of the sign has in his own mind. (Augustine, *On Christian Doctrine*: Chapters 2.1 and 2.2)

I say vocal words are signs subordinated to mental concepts or contents. By this I do not mean that if the word 'sign' is taken in its proper meaning, spoken words are properly and primarily signs of mental concepts; I rather mean that words are applied in order to signify the very same things which are signified by mental concepts. Hence the concept signifies something primarily and naturally, whilst the word signifies the same thing secondarily. . . . This is what is meant by the Philosopher when he says 'Words are signs of the impressions in the soul'. Boethius also has the same in mind when he says that words signify concepts. . . . A concept or mental impression signifies naturally whatever it does signify; a spoken or written term, on the other hand, does not signify anything except by free convention. (Ockham, *Summa Logicae I*: §1)

Three centuries later, the modern period of philosophy began with similar statements by Descartes in the *Meditations* (1641) and *Replies to Objections* (1641), Hobbes in the *Logic* (1655), and Arnauld in the *Port Royal Grammar* (1660) and *Port Royal Logic* (1662). Descartes introduced the term

"idea" in this context, which became firmly entrenched through the enormous influence of Locke's *An Essay Concerning Human Understanding* (1690).

> Man, though he have great variety of thoughts, and such from which others as well as himself might receive profit and delight; yet they are all within his own breast, invisible and hidden from others, nor can of themselves be made to appear. The comfort and advantage of society not being to be had without communication of thoughts, it was necessary that man should find out some external sensible signs, whereof those invisible ideas, which his thoughts are made up of, might be made known to others. For this purpose nothing was so fit, either for plenty or quickness, as those articulate sounds, which with so much ease and variety he found himself able to make. Thus we may conceive how *words*, which were by nature so well adapted to that purpose, came to be made use of by men as the signs of their ideas; not by any natural connexion that there is between particular articulate sounds and certain ideas, for then there would be but one language amongst all men; but by a voluntary imposition, whereby such a sound is made arbitrarily the mark of such an idea. The use, then, of words, is to be sensible marks of ideas; and the ideas they stand for are their proper and immediate signification. (Locke 1690: §3.2.1)

> Words, by long and familiar use, as has been said, come to excite in men certain ideas so constantly and readily, that they are apt to suppose a natural connexion between them. But . . . every man has so inviolable a liberty to make words stand for what ideas he pleases, that no one hath the power to make others have the same ideas in their minds that he has, when they use the same words that he does. And therefore the great Augustus himself, in the possession of that power which ruled the world, acknowledged he could not make a new Latin word: which was as much as to say, that he could not arbitrarily appoint what idea any sound should be a sign of, in the mouths and common language of his subjects. It is true, common use, by a tacit consent, appropriates certain sounds to certain ideas in all languages, which so far limits the signification of that sound, that unless a man applies it to the same idea, he does not speak properly. . . . (Locke 1690: §3.2.4)

Locke's views on the signification of ideas were repeated with very little variation or amplification for the next three centuries, principally by those who thought that they had found in the principles of association formulated by both Locke and Aristotle the fundamental laws of all mental phenomena.[1] Even those who rejected associationism accepted

1 See Condillac 1746; Hartley 1749: Chapter 1.3; J. Mill 1829, Chapter 4; Bentham 1816, 1843; Bain 1855: §67–8; and Titchener 1914: 214.

a basically Lockean view of language.[2] One of the few new ideas was the late nineteenth-century distinction between sense and reference, which led to Frege's thesis that the sense of a sentence is a thought, the sense of a predicate a concept.[3] Frege's identification of senses with thoughts is subject to objection, but its ability to account for the distinction between sense and reference is a major strength. J. S. Mill's similar but older distinction between connotation and denotation led to a significant competitor to the ideational theory: the view that the meaning of a word is its connotation, the property or set of properties it expresses.[4]

As this brief history indicates, the expression theory underwent little development between the third century B.C. and the first half of the twentieth century. Critics have been more inventive, developing a multitude of objections. Much of the classical criticism has centered around the notion of an idea. The expression theory is primarily, though not exclusively, an ideational theory. Ideational theorists tended to use the term "idea" inconsistently, and many definitions picked out classes of entities that did not correlate well with meanings. The term "idea" became enmeshed in wildly implausible theories such as idealism, associationism, and sensationalism. The philosophical pendulum swung hard in the opposite direction, producing the verification theory of the Vienna Circle, according to which meaning consists in verification conditions. Bertrand Russell and Wittgenstein in his early writings advocated the referential theory, identifying meaning with reference. The later Wittgensteinian dictum that "meaning is use" resonated with the behaviorist movement that was sweeping philosophy as well as psychology. The evident failure of behaviorist analyses, which was as great for semantic terms as for psychological terms, led Quine and his followers to reject as meaningless all talk of meaning as opposed to reference. The rapid progress of modern formal logic rewarded work on reference, and underscored its relative tractability.

2 See Leibniz 1709: 3.1–3.2; Reid 1764: §4.2, §5.3, §6.24; Reid 1785: 394, 477, 496–7; Brentano 1874: 198; James 1890: 427; Frege 1892a: 43; 1892b; 1918: 4–5; Husserl 1900: Investigation I; Meinong 1910: xiv–xv, 24–5, 34–6.

3 Frege 1892a: 43; 1892b; 1918: 4–5. See also Husserl 1900: Investigation I, Chapter 1, §12; Kneale & Kneale 1962: 493ff.

4 J. S. Mill 1843: §1.2.5; §1.5.2; §1.5.4. According to Kneale & Kneale 1962: 318, this distinction between "comprehension" and "extension" was first introduced by Arnauld in the *Port Royal Logic*. Hamilton introduced "intension" for "comprehension." Ockham's distinction between secondary and primary signification would seem to be an early predecessor. See Loux 1974: 6–7; Freddoso 1980: 4–5; Ockham, *Summa Logicae I*. Formally, however, Ockham had no use for the abstract objects or universals that connotative terms appear to signify secondarily.

General arguments for eliminative materialism gained currency, leading some to reject mentalistic theories of language altogether. By the mid-1960s, the ideational theory of meaning was as dead as idealism in metaphysics.

I will not in this work attempt to refute skepticism about the existence of mental phenomena. I will take it for granted that people do have beliefs, desires, and thoughts. Let the skeptics be taken at their word that they do not really believe or mean what they are saying, and have not thought the matter through! Seriously, I believe that we have direct introspective evidence for the existence of beliefs, desires, and thoughts, and indirect evidence based on the ability of psychological hypotheses to explain and predict human behavior, including but not limited to verbal behavior. There is an impressive and rapidly expanding literature on the neurophysiological basis of psychological phenomena. The fact that there are no serious competitors to explanations of behavior in terms of mental states has been argued forcefully by Chomsky, Putnam, Fodor, and others, and many results from the burgeoning field of cognitive psychology demonstrate the power of the framework. The case for the predictive value of psychological hypotheses has not been made as thoroughly, so I will make one observation. The triumph of the Apollo moon missions was rightly attributed to the remarkable predictive power of physical theory, which enabled scientists to calculate in advance the exact path the capsule would take, the amount of fuel needed to return the ship to Earth, and so on. It is seldom observed that the success of the mission depended equally critically on the scientists' ability to predict the behavior of the astronauts manning the spacecraft. These predictions were based not on the laws of physics or neurophysiology, but on the known psychological states of the astronauts and the principles by which such states lead to behavior. Mission control knew, for example, that the astronauts wanted to get to the Moon and return safely, that they believed a number of specific actions were necessary to achieve that goal, that the astronauts would think of the necessary actions at the appropriate times, and that the actions would be performed at the right times as a result. The predictive power of psychology is astonishing when you think about it.

Skeptics like Churchland (1981) myopically focus on the unexplained and the unpredictable. Every advance in scientific understanding raises more questions than it answers. Churchland also makes much of the fact that psychology has advanced comparatively little in three thousand years, concluding that it is a "stagnant research paradigm." But the relative stagnation has some obvious explanations: the mind is enormously complex;

scholars did not think to study psychology scientifically until around 1850, and when they did a number of false and unwarranted dogmas – principally sensationalism and behaviorism – obstructed research for the next hundred and twenty years. We will in this work patiently untangle conceptual confusions and theoretical dogmas that impede understanding of meaning and thought even today. Despite these obstacles, today's introductory psychology and semantics texts represent a vast improvement on *De Anima* and *De Interpretatione* in any number of ways. Progress has not been as great as in physics, to be sure, but neither has there been stagnation.

What I will try to do, at some length, is to clarify the sense of thought and ideation for which the expression theory of meaning holds true. I will focus on thinking as a propositional attitude distinct from believing and desiring, and will define ideas as thoughts or parts of thoughts. This will produce a theory similar in many respects to the "language of thought hypothesis," which, I shall argue, cannot be taken seriously on its most common interpretations. Thoughts, on my view, are structured events, a particular kind of mental representation. They are similar in many ways to sentences, but are fundamentally different, and more fundamental. Thoughts are propositions in the sense in which belief and desire are propositional attitudes. We think when thoughts occur to us. All other propositional attitudes are different relations to thoughts. Thoughts have constituent structure in a literal sense that beliefs and desires do not. Since thoughts are readily introspectible, the failure to grant thought its proper place in psychology has deprived cognitive scientists of a large and fascinating body of data as well as crucial theoretical resources.

Ideational theorists have traditionally held that the meaning of a word is an idea, and that the sense of a sentence is a thought (§21.1). This identification is untenable, and I avoid it. Meanings are properties of words, their expressing ideas or other mental states. Another powerful objection to the ideational theory is that it merely defines one semantic notion in terms either of itself or of another notion equally in need of analysis. One line of thought in this direction depends on failing to perceive the significant differences between the meanings of words and the contents of ideas (Chapter 22). Another starts reasonably from physicalism, but then insists groundlessly that the semantic cannot be defined in psychological terms unless those are first defined physically (Chapter 23). A legitimate objection is based on the observation that we have gained little by explaining what a word *means* in terms of what a word *expresses*, unless we can explain expression independent of meaning. Indeed, it is natural to suspect that "expressing the idea *man*" just means "meaning *man*." This suspicion leads

naturally to the conclusion that idea-talk, to the extent that it is legitimate, is just a *façon de parler* to be explained away in terms of meaning-talk, with the consequence that the ideational theory cannot possibly tell us what meaning is. The attempt of Hobbes and Locke to define "mark" was a step in the right direction, but hardly went far enough to be satisfying. The problem was not perceived, let alone addressed, by their followers. The solution, I believe, can be found by developing the work of H. P. Grice,[5] who attacked the problem of meaning from a completely different direction.

§1.2 THE GRICEAN PROGRAM

Grice had a bold and original vision, one that has attracted scholars well beyond the bounds of philosophy. What words mean, Grice observed, is determined in some way by what speakers mean by them. What speakers mean is determined in some way by their intentions. The central subject of linguistics, it follows, is fundamentally a matter of psychology. While Grice's initial attempt to specify how word meaning is related to speaker meaning was crude and unsuccessful, his attempt to define speaker meaning in terms of intention has found a wide following, and stimulated a large body of research.

To mean something, Grice said, is to act with the intention of producing a certain response in one's audience by means of recognition of intention. Many thought that this analysis was basically right, and offered minor variations to handle a few tricky cases. Others observed that with slight adjustments, the Gricean condition could be used equally well to define the related notions of expressing and referring. Unfortunately, a broad and diverse body of familiar facts seems to show quite clearly that the Gricean analysis of meaning, expressing, and referring is fundamentally flawed. The principal error, I shall argue, is its emphasis on audience-directed intentions. This very feature, however, makes the Gricean condition a natural candidate for the analysis of informing, telling, and communicating. Indeed, almost as many philosophers have used the Gricean condition to define communication as have advocated it for meaning. Grice and his closest followers wrongly assumed, I shall argue, that meaning is the attempt to communicate.

I refer to expressing, referring, and communicating as *semantic acts*. They comprise a special class of illocutionary speech acts, distinguished in part by their fundamentality. "Speech act theory" has come to be

5 See Grice 1957, 1968, 1969a, 1982, 1986, 1989.

understood as the study of asserting, ordering, questioning, requesting, promising, apologizing, begging, and other similar actions. Much attention has been devoted to their classification and definition. All entail the performance of what I am calling semantic acts. It is impossible to assert something without expressing a belief. You cannot ask someone to do something without referring to and communicating with her. By contrast, none of the higher-order illocutionary acts are entailed by the semantic acts. I can express the belief that someone is asleep without asserting that he is. I can express a desire for you to leave without ordering or asking you to leave.

I shall define communication in terms of meaning, meaning and referring in terms of expression, and expression in terms of intention (Part I). I shall thus be carrying out part of Grice's program. But the intention that I specify differs markedly from the Gricean intention. We need to distinguish clearly between Grice's general program, and the specific implementation that he proposed. In place of the intention to produce certain responses in an audience, I substitute the intention to produce an indication that one has certain mental states. Since indication is a close relative of what Grice called natural meaning or signification, my account is more Aristotelian or Lockean. I hope to make it clear that my analysis accommodates simply and naturally the whole dizzying array of facts that are problematic and must be explained away on competing theories. The Gricean analysis, for example, has trouble with the familiar fact that people often talk to babies, and mean something when they do, despite having no intention to produce a belief in them. Hand waving, bullet biting, and other desperate measures are unnecessary on my account: someone talking to a baby is still expressing thoughts and beliefs, and does intend to provide an indication that he has them. The stultification and defeatism characterizing recent work on semantic acts is a product, I submit, of an irrational fixation on the specifics of Grice's proposal.

While semantic acts are the central focus of Part I, our attention will not be confined to them. Complete understanding of the act of communicating, for example, requires comparing it to informing and telling, which are higher-order illocutionary acts. Since understanding a speaker requires grasping what he or she means, we will implicitly be shedding light on the nature of understanding. We will not, however, be investigating how we understand words, or how we produce meaningful speech, which are empirical problems for cognitive science. Our general goal is to shed light on meaning, expression, communication, and reference by showing the exact location of these concepts in our larger conceptual scheme. In the

process, we will distinguish two different kinds of communication and four different kinds of speaker meaning. Given the importance to human beings of the general activity of conveying ideas, it should not be surprising that we have a complex system of concepts specifying different but closely related ways of doing so. My goal is to delineate the structure of this system.

Toward the same end, I will also attempt to explain how word meaning is related to speaker meaning, thus tackling the other part of the Gricean program (Part II). This will involve defining what it is for a speaker to use a language, and for a language to be a living language. The theory to be advanced is simultaneously a use theory, an ideational theory, an intentionalistic theory, and a conventionalist theory. The relation between truth and meaning will be duly explained. A language, on my view, is a system for the expression of ideas and other mental states. It is a living language only if it is used conventionally by a group of speakers for the purpose of communication. What words mean in a living language is dependent on the conventions in that group governing what speakers mean by their words. The referential properties of words are those of the ideas they express. I hope to show, in short, that generative semantics and sociolinguistic pragmatics are complementary rather than competing approaches to the study of language. To round out the study, we will briefly examine meaning in artificial languages and idiolects. Meaning here is established by stipulation and individual practice, respectively, rather than by convention.

One problem that Grice did not see arises from the relativity of word meaning to languages. "Rot" means "red" in German, "decay" in English. If we say that "rot" means "red" in German because it is conventional for German speakers to use "rot" to mean "red," do we not run in a circle? For what is it to speak German except to use words to mean what they mean in German? How can we pick out the conventions that determine what words mean in German without identifying them as the conventions to use words to mean what they mean in German? The answer will lie partly in the self-perpetuating character of conventions, in virtue of which usage today evolves from prior usage, and partly in our ability to identify new languages without knowing what words mean in those languages.

Grice and his followers attempted to define word meaning in terms of one kind of speaker meaning: meaning *that* p by uttering e, which involves the expression of belief. This led, among other things, to difficulties with linguistic units below the level of the sentence, and to the problematic doctrine that word meaning must be defined in terms of sentence meaning. I avoid these difficulties by focusing on another kind of

speaker meaning: meaning "m" by expression e, which involves the direct expression of thoughts or ideas. In another respect, then, my neo-Gricean account is more Lockean or Aristotelian than Grice's own. Our ability to define word meaning independent of sentence meaning will enable us to account for the compositionality and productivity of meaning. The meaning of a sentence is determined recursively by the conventions pairing word structures with idea structures, and by the basic conventions pairing the words in the sentence with ideas.

One of my main subthemes is that thought is as fundamental and important a concept of psychology as belief or desire. Accordingly, I will devote considerable attention to distinguishing thought from belief, developing the notion of ideas as thought parts (Part III). I will then rebut objections to ideational theories of meaning (Part IV). Given that our goal is to understand the nature of meaning, the "*de dicto*" attitudes will generally be more important for us than "*de re*" attitudes (see §6.2). Sentences ascribing propositional attitudes generally have transparent as well as opaque interpretations. The opaque interpretation will be our default.

While names have seemed especially problematic, they are easily accommodated by the expression theory once we abandon the widely accepted but groundless doctrine that all basic ideas are general or descriptive. Indexicals are more complicated than names, but can also be handled by the expression theory. Indexicals express a special type of thought-part that links with perceptions and other mental events, whose reference becomes the indexicals' reference. I will develop the expression theory for these two domains in my forthcoming *Nondescriptive Meaning and Reference: Names, Indexicals, and Other Special Cases*. The other cases will include interjections, syncategorematic terms, conventional implicatures, and pejorative terms.

§1.3 SYSTEMATIZATION

This work is part of a larger project, which seeks to increase our understanding of psychology by systematizing its elements.[6] Psychology, as I understand it, is the study of belief, desire, thought, intention, decision, reasoning, inference, fear, hope, joy, sorrow, pain, imagination, itches, tingles, sensations, and all of the other states and processes either available to introspection or closely related to those that are. Psychologists are

6 See the brief history of my *Belief, Desire, and Thought* in the Preface. Parts of the system have been published in Davis 1981a, 1981b, 1982, 1984a, 1984b, 1987, 1988a, 1988b.

empirical scientists specially trained in the art of observation, experimentation, and theory construction. In addition to the evidence available only to psychologists, there is a large body of common psychological knowledge. Some of this common knowledge is a priori, such as the knowledge that belief is different from desire and incompatible with disbelief. Some of it is a posteriori but nevertheless available to intelligent observers without specialized training, such as the knowledge that people who want to do something do not always do it. Part of it is linguistic, based on our knowledge of the language that we use to talk about psychological phenomena. Much of this knowledge is learned early as a part of normal maturation, rather than through formal study or schooling. By the "elements" of psychology, I mean the general principles that are either part of, or can be based on, this common knowledge. Note well that principles based on common knowledge need not themselves be common knowledge.

One way to increase our understanding of a subject is to systematize what is known about it. One way to systematize a body of knowledge is to organize it into a deductive system. In such a system, some terms are taken as primitive and others are defined. Some principles are taken as postulates and others are derived as theorems. Other things being equal, the greater the completeness and economy of the system – the greater the proportion of the knowledge incorporated, and the fewer the number of primitive terms and principles – the better the systematization. To the extent that it increases the integration of our knowledge, systematization increases our understanding. The project of systematizing the elements of psychology thus involves generalizing and integrating the body of common psychological knowledge. The goal is an increased understanding of psychological phenomena. The resulting system provides a framework for formulating the results of specialized empirical research, one that is general enough not to constrain it.

Expressing, referring, and communicating are psychological acts. Defining semantic acts in terms of intentions therefore contributes directly to the systematization of psychology. I believe that intention itself can be defined in terms of belief and desire. But that part of the systematization will not be presented here. Many philosophers are interested in the Gricean program because they believe that belief and desire will ultimately be identified with neurophysiological states, holding out the prospect of reducing semantics ultimately to physics. While I myself believe that the progress of science indicates that belief and desire are neurophysiological states, I will not be concerned to argue that they are. Resolution of the mind-body problem is not required to settle any of the issues we shall

confront. We will similarly remain neutral concerning Locke's empiricist thesis that all ideas are derived from experience by abstraction or definition, and its nativist rival.

Much of the literature in cognitive psychology is devoted to describing mental phenomena in metaphorical terms derived from computer science. An air of respectability is thereby created, based on the success of that field. The practice is much like that of Hartley (1749), who described the association of ideas in terms of vibrations in the neural aether. I believe that such metaphors are as blinding as they are illuminating. They also divert our attention to lesser tasks. We could easily spend a lot of time, for example, in a fruitless debate over whether thinking really counts as the "processing" of information, or as "operating on" a representation. I prefer the direct and literal description of the phenomena I am interested in. I am confident that anything that is precisely described can be represented digitally with a high degree of accuracy.

§1.4 ANALYSES

Most of the principles to be presented here are either definitions, or theorems following from the definitions. To define a given sense of a term is to state necessary and sufficient conditions for its application in that sense. The definition of speaker meaning, therefore, will take the form "S means that p if and only if S directly expresses the belief that p." The cumbersome "if and only if" will be abbreviated "iff." Definitions are often abbreviated further by putting them in the subject-predicate form "S is P." But when intended as definitions rather than mere predications, sentences of this form must imply that something is S if and only if it is P. An example is provided by the definition of convention offered in Chapter 9: "A convention is a regularity that is socially useful, self-perpetuating, and arbitrary," which implies that something is a convention iff it is such a regularity. Correct definitions must be neither too broad (meaning that the stated conditions are not in fact sufficient) nor too narrow (meaning that the stated conditions are not in fact necessary). That is, there must be no counterexamples.

There are many kinds of definition, serving different purposes.[7] Since our goal is to increase our understanding of psychological and linguistic phenomena, we will attempt to provide definitions that tell us *what it is* for something to be what we are trying to define. We will try to explain what

7 For an introductory survey, see Davis 1986: §10.4.

it is for a word to have meaning, and more specifically what makes a word mean "red," for example. We will accordingly look for defining conditions that are *essential*. Such definitions will be called *analyses*, without suggesting that the definiens and definiendum have to be synonymous or logically equivalent. It is true, for example, that something is a penny if, and only if, it is the U.S. coin with the lowest value. But this equivalence does not count as an analysis in our sense, since it does not tell us what it is for a coin to be a penny. Having the lowest value is not what makes a coin a penny. It is not essential to being a penny. Hence it is quite possible that the United States will someday introduce a coin lower in value than the penny, or (more likely) discontinue the penny, making the nickel the lowest-value coin. While analyses need not be *logically* necessary, they do need to hold in all genuinely possible cases. The conditions provided need to be "nomically" or "metaphysically" necessary and sufficient. Hence in determining whether a definition is too broad or too narrow, we have to consider hypothetical cases as well as actual ones.

The term "analysis" is typically used to mean a definition that is analytically true, that is, a definition in which the definiens and definiendum have the same meaning. "A bachelor is an unmarried man" is the classic analysis in this strict sense. We are using the term "analysis" without the requirement that the definitions be analytic or even logically true, because these properties are not necessary for a definition to tell us what it is for something to be what we are trying to define. "Water is H_2O" tells us in a most informative way what water is, even though the definition is synthetic and logically contingent.

In addition to having no counterexamples, analyses must be individually as well as collectively noncircular. It is necessarily true that something is a cat iff it is either a white cat or a nonwhite cat. But this necessary and sufficient condition does not tell us what a cat is, or explain what makes something a cat. The best way of increasing our understanding of meaning is to define it in terms that are not themselves defined in terms of meaning, and only a noncircular definition can tell us what meaning is. We will therefore provide careful replies to objections that ideational theories of meaning in particular, and mentalistic theories generally, are circular in various ways (Chapter 22). And as indicated earlier, we will be very sensitive to the concern that we have gained little by defining meaning in terms of expression unless we can define expression independent of meaning.

The definitions that I present are generally intended not as stipulations, but as statements that are true when the semantic terms to be defined are

interpreted in their conventional senses. This does not mean that my purpose is simply to describe "ordinary language." Rather, my purpose is to describe those acts we are normally talking about when we use the terms "mean," "express," "refer," and "communicate" (or their equivalents in other languages). We want to know what it is for those acts to be performed. We could, and sometimes do, use these terms to talk about other matters. And there are plenty of other things people do that are worthy of study. But the fact remains that we all have a deep and abiding interest in what speakers mean by their words, what beliefs, thoughts, and desires they are expressing, what objects they are referring to, who they are communicating with, and so on. Our response to a speaker generally depends on our determinations in these matters, and the appropriateness of our response depends on their accuracy.

The attempt to define the ordinary sense of semantic terms is sometimes thought to be misguided. As Devitt (1981: 88) put it:

> Philosophy is an area of knowledge, like others, concerned with theorizing about the world. Our concern here is to produce a *theory* about linguistic phenomena. The correct theory in semantics is no more likely to be discovered by examining ordinary semantic terms than is the correct theory in physics to be discovered by examining ordinary physical terms.

I too am concerned to produce a theory about linguistic phenomena. But there are lots of linguistic phenomena. Whenever a man opens his mouth, he performs a dozen different linguistic acts. He moves his articulatory apparatus in certain ways, produces certain speech sounds, utters certain words, means something by those words, refers to things, expresses certain thoughts, and implies something; he makes a statement, issues an order, or asks a question; he follows or violates linguistic rules; he communicates or fails to communicate with an audience, informs, bores, or annoys them; and so on and on. I need to tell you which of all those linguistic acts I am going to theorize about. Since I am writing in English, I am using the English words that are conventionally used to express the semantic acts I am interested in. The fact that my subject matter is expressed by words used every day by billions of people, which are learned early in life by all who develop normally, is an indication that my interest in the subject is not a personal quirk, nor the product of a fad. A linguistic or psychological theory that ignores these acts may be correct, but it cannot be complete. It is generally easy to produce a true theory by excluding difficult phenomena from its scope, or by changing the subject. Our goal is not just truth, but the whole truth.

Nothing I have said implies that in addition to meaning, referring, and the like, the speaker is not also doing many important things for which there are no conventional expressions. I am even willing to grant, as a logical possibility, that linguists and psychologists might someday show that meaning, expression, and communication (in the ordinary senses of these terms) are not actions of fundamental linguistic importance, just as physicists have shown that the color of an object is not of fundamental physical importance. But given what is currently known, this possibility seems extremely remote. It may also turn out that these terms are too vague to be scientifically useful, or that they classify together actions that have fundamentally different explanations. But that will not be established until we have a much better understanding of semantic terminology than we currently possess. I am quite confident that a theory of meaning that does not talk about meaning must fail.

I distinguish meaning from referential properties like truth and denotation in the usual way (e.g., §8.3). It may be wondered why we should study meaning when referential properties are so important, and when logicians and formal semanticists studying them have made so much progress. The motivation is simple: meaning is *also* important – indeed, it is one of the determinants of referential properties. Meaning is also more closely related to psychology, which is my primary interest. A final motivation for studying meaning is that it is *not* well understood. Scholars in all fields, and particularly philosophy, seek to understand what is not yet understood. It is laziness to avoid something because it is difficult, and cowardice to do so because the risk of failure is great.

As for Devitt's reference to physics, it should be recalled first that "truth" and "denotation" are also "ordinary" semantic terms. Second, one of the elements of the scientific revolution leading to the success of Newtonian mechanics was a sustained examination of the quite ordinary physical concept of motion, resulting in its analysis in terms of the ordinary concepts of "change," "place," and "time." That analysis and its application built on similar results achieved a millennium earlier for the ordinary concepts of point, line, plane, triangle, circle, and the whole Euclidean system. It would be foolhardy to predict the same measure of success from a study like this one. But it would be equally unwise at this stage in the investigation of language to propose a theory of linguistic phenomena without a careful examination of the concept of meaning.

Schiffer (1987a: 248) cites the "dismal history of analysis" as evidence that speaker meaning cannot be defined in terms of intention or anything else – that there is no correct, interesting, noncircular completion for

"S means that p iff ____."[8] I shall argue that there is more than one. But let us suppose that Schiffer is right. Would that make any attempt such as mine to improve on extant analyses a mere exercise in futility, a pointless waste of time? Certainly not. Schiffer's schema, like those to be presented here, has an implicit universal quantifier. Universal generalizations are not the only useful or important generalizations. "Nearly all people die before they are one hundred years old" is a very important fact about human beings even though it allows for exceptions. It is not obviously less informative than the completely universal but temporally indefinite "All people die eventually." "Almost all human beings are able to learn a language" is a fundamentally important fact about human beings even though it is not a universal generalization. It would represent a considerable intellectual achievement if we could truly and without circularity complete a schema of the form "*With few if any exceptions*, S means that p iff S has such-and-such intentions." This would not count as a complete conceptual analysis or definitional reduction, nor as a complete theory of how speaker meaning depends on intention. But it would be highly informative nonetheless, and would tell us *at least roughly* what speaker meaning is, and how it depends on intention. Similarly, it seems clear that what words mean in natural languages depends in some way on what speakers of those languages use them to mean. It may not be humanly possible to state the nature of this dependence in its full generality. But it would be nice to find a rule that held in at least some cases. The greater the number and variety of cases for which the formulation holds, the better.

The objection to previous analyses in this light is not that they are not quite universally true. The problem is that the exceptions are many and various, and occur in the most familiar of cases. Better analyses are surely possible.

Expression theorists from Aristotle to Frege barely scratched the surface of the relationship between language and thought. It is remarkable how little progress was made over such a long period. This can perhaps be explained by the fact that expression theorists were primarily concerned with other areas: Aristotle and Frege with logic; Locke with epistemology; the association psychologists with the reduction of mental phenomena to sensation; functionalists with physicalism, and so on. We shall keep our attention firmly fixed on meaning, expression, and thought.

8 See also Fodor 1975: 124–56; 1981: Chapter 10; 1987: 161; Dummett 1975: 97–8; McDowell 1980: 124; and Stich 1983: 76–8. Compare and contrast Avramides 1989: §1.3; 1997: §4.

Part One

Semantic Acts and Intentions

2

Speaker Meaning

The term "meaning" expresses a close-knit family of concepts. In order to properly identify our subject, and to prevent equivocation, the concepts must be carefully identified and differentiated. In this chapter, we will distinguish speaker meaning from other types, and then discriminate several senses in which a speaker can mean something, defining them all in terms of expression, along with the related concept of speaker implication. Of particular importance for later developments will be the type of speaker meaning involving the expression of thoughts or ideas, in a sense of thought distinct from belief. Competing definitions of speaker meaning will be discussed in Chapter 4, after expression has been defined in Chapter 3. The reader might find it helpful along the way to examine Figure 2.1 at the end of this chapter, which presents all of the different senses of meaning we shall discuss in a classification tree.

§2.1 SPEAKER, WORD, AND EVIDENTIAL SENSES

A glance through any dictionary will show that the most important and commonly used terms generally have a large set of meanings, some closely related and others quite distinct. "Means" is no exception. Since this is a book on meaning, it will be essential for us to distinguish the senses of "means" even more finely and more sharply than is profitable in a dictionary. The three most important senses for us are represented in the following sentences.

(1) Boulders mean glacial activity. (*Evidential Meaning*)
(2) "Boulder" means "large rounded stone block." (*Word Meaning*)
(3) By "boulder," S means "kilo of cocaine." (*Speaker Meaning*)

Sentence (1) says that boulders *indicate* and *provide evidence of* glacial activity, which is true because the causal connection between boulders and glacial activity enables us to infer the latter from the former. Since "indicate" and "provide evidence of" are close synonyms of "mean" in examples like (1), I refer to this sense of meaning as *indicial* or *evidential meaning*. "Indicate" and "provide evidence of" are not even rough synonyms of "mean" in (2) or (3). They do not tell us what we can infer from the word "boulder" or the speaker, and do not imply that there is anything like a causal relationship between the word "boulder" and the two phrases. Boulders "*carry* information about" glacial activity, and *utterances* of the word "boulder" carry information about a variety of things – including but not limited to occurrences of the concept "large rounded stone block." But the word itself does not carry information about anything. The word is "*used to convey* information about" stone blocks; but such blocks are not used to convey information about glacial activity or anything else. Because the word "boulder" is ambiguous (large marble, type of clay deposit, city in Colorado) and could also be used to mean something unconventional (kilo of cocaine), sentence (2) does not even entail that a majority of occurrences of "boulder" indicate anything related to large rounded stone blocks or the concept thereof.

Following traditions established by Grice and Pierce, I use *word* or *symbolic meaning* for the sense of meaning illustrated by (2), since in this sense the verb "means" takes a subject referring to things like words or symbols. I use *speaker* or *agentive meaning*, similarly, for the sense of meaning illustrated by (3), since in this sense the verb "means" takes a subject referring to an individual who has spoken, or who has done something similar such as writing or gesturing. Thus (3) entails that the object to which "means" is ascribed did something; (2), by contrast, does not: words are not agents. Sentence (2) entails that the subject of meaning has meaning and is meaningful; sentence (3) does not. Hence only in (2) is "means" interchangeable with "has the meaning." Only when "means" applies to a word can it be qualified by an adverbial phrase specifying a language. "'Rot' means 'red' in German" makes sense, but "Steve means 'red' in German" does not.

Despite their differences, there are many important similarities and interconnections among the three senses of meaning. Indeed, we shall eventually define word meaning in terms of speaker meaning, and speaker meaning in terms of evidential meaning.

The discipline of semantics is concerned with both word and speaker meaning, but not with evidential meaning. Consequently, I will classify

word and speaker meaning as forms of *semantic meaning* when distinguishing both from evidential meaning. Both forms of semantic meaning depend on intelligent action and intention. Neither (2) nor (3) would be true unless someone had used the word "boulder" for certain purposes. Evidential meaning, by contrast, does not generally depend on intention or intelligent action of any kind. Sentence (1) could be true even if human beings or other intelligent agents had never evolved on the planet. People are required to interpret or use the evidence provided, but indications need not be interpreted in order to exist.

It is natural to suspect that word meaning is the more fundamental semantic notion, and that "S means μ" is simply short for "S used a word that means μ" or "S intended to use a word meaning μ." I call this the *naive analysis* of speaker meaning, because a little reflection suffices to show that it is a mistake.

(4) S means μ by e iff S used e and (or because) e means μ.[1]

Recognizing how badly mistaken the naive analysis is will be essential to appreciating the developments that follow. Codes, nonce words, technical stipulations, figures of speech (metaphor, simile, irony, hyperbole, metonymy, etc.), ellipses, shorthand expressions, and linguistic mistakes, not to mention pronouns and ambiguities, constitute a large and vital part of language. They are cases in which what a word means differs from what a particular speaker means by it. Thus (3) may be true as well as (2) because S is speaking in code. Similarly, if Bill makes a common mistake and says "Dr. Johnson is an entomologist" meaning that she is an etymologist, then Bill means "etymologist" by "entomologist" even though "entomologist" does not mean "etymologist." A boy describing something as "a mile long" may have been speaking ironically (meaning "very short") or hyperbolically (meaning "very long"). Finally, from the fact that the ambiguous word "broom" has the meaning *Cytisus scoparius*, it does not follow that this is what Aunt Hilda meant by it. This may be true even if one of the speaker's reasons for choosing that particular word was that it had this additional meaning.

I will argue that speaker meaning is the more fundamental semantic notion. For a brief characterization, we may say that "'Boulder' means 'large rounded stone block'" is roughly equivalent to "Speakers conventionally

1 Cf. Black 1972: §VII; Dummett 1973: 149; McKinsey 1978: 191–2; Harrison 1980: 193; Skulsky 1986: 593; Pettit 1987: 729; Laurence 1996: 285–6, 292–6. Contrast H. H. Clark (1983, 1993), who describes "the ubiquity of nonce sense."

mean 'large rounded stone block' by 'boulder.'" Chapters 8–11 will be devoted to producing an accurate and general definition along these lines.[2]

In sentences (1) to (3), "means" expresses different forms of *signification*. "Signifies" can replace "means" in each case, and has the same three senses. "Signifies" has the convenient correlative "signifier," which "means" lacks. Nonagentive signifiers may be classified as *signs, symptoms, marks, signals, symbols, gestures, words*, and so forth. Evidential signifiers are signs, symptoms, or marks *of* what they signify, while symbolic signifiers are signs, marks, or symbols *for* what they signify. Thus boulders are signs of, not for, glacial activity; and "+" is a sign for, not of, addition. Evidential signifiers are Peirce's "indexes," symbolic signifiers his "symbols."[3] We will not be concerned with the subtle distinctions among semantic signifiers marked by the terms "sign," "signal," "symbol," "word," and so on. For we are interested in meaning, and all of them mean what they do in the same sense. For example, "boulder" is properly characterized as a word, but not as a sign. Its translation into American Sign Language, on the other hand, is properly characterized as a sign but not as a word. Nevertheless, both are signifiers, and both mean "large rounded stone block." We will use *expression* neutrally for the broad class of nonagentive signifiers.

Grice (1957) used "natural meaning" for evidential meaning, and classified both word and speaker meaning as "nonnatural meaning." Grice's usage, however, is confusing in light of the venerable distinction between *natural* and *conventional signs*. Hobbes introduced the distinction as follows.

> Signs however are customarily called *the antecedents of consequences and the consequences of antecedents, since we generally experience them in a similar way preceding or following one another in a similar fashion*. For example, a dense cloud is a sign of consequent rain and rain a sign of an antecedent cloud, for the reason that we know from experience that there is rarely a dense cloud without consequent rain, and never rain without an antecedent cloud. Of signs, however, some are natural of which type we have just discussed an example. Others are conventional, namely, those which are applied of our own accord; of this type are: a bush hung for signifying that wine is for sale, a stone for signifying the boundaries of a field,

2 In §7.9, we will distinguish what a word means *simpliciter* from what it means *on a given occasion*.

3 See Peirce 1931–5: 2.247, 2.248, 2.307; Burks 1949; and Alston 1964a: 55–6. Cf. Meinong 1910: 23 and Husserl 1900: 269–73. We will not be concerned with Peirce's "icons," illustrated by schematic diagrams and pictures, because they cannot be said to mean or signify what they represent. This is not to deny that icons often have symbols as parts.

and human vocal sounds connected in a certain way for signifying the thoughts and motions of the mind. (Hobbes 1655: §2.3)[4]

Hobbes's distinction applies only to evidential signs. Indeed, evidential signification is the type of meaning that Hobbes was mainly concerned with.[5] So even Hobbes's conventional signs mean what they do in Grice's "natural" sense. The fact that John said "It is raining" may mean (i.e., indicate) that John is thinking the thought that it is raining. But John's utterance is a conventional sign of the thought rather than a natural sign. For English sentences are artificial, and indicate what they do in virtue of certain conventions. Note well that while utterances of "It is raining" signify *evidentially* the thought that it is raining, the sentence signifies *symbolically* that it is raining. Convention is involved in both relationships, but in different ways. The same may hold when the signified and signifier are identical, as when a turn signal signifies a turn both evidentially and symbolically. We will have more to say about this later.

While important, Hobbes's distinction is neither exhaustive nor free from vagueness. When speakers use codes, for example, we have non-natural signs that are not happily described as conventional. John's use of "boulder" may indicate that he is thinking of cocaine, for example, but not in virtue of either nature or convention. And John's conventional uses of "boulder" indicate that he knows English, that the language areas of his brain have not been ablated, and so on. But as a sign of his knowledge of English or of the condition of his brain, John's utterances are not comfortably classified as either conventional or natural. Since distinctions among evidential signs will not be important for us, we will not try to clarify or extend Hobbes's distinction.

In other senses, "mean" does not express a form of signification. Whereas "means" can be replaced by "signifies" in sentences (1)–(3), replacement by "signifies" is not possible in (5) or (6).

4 See also Augustine, *On Christian Doctrine*: Chapters 2.1–2.2; Arnauld 1662: 47. Reid 1764: §4.2, §5.3 draws a similar (if slightly different) distinction between natural and artificial signs, which turns on whether the sign's meaning is "fixed by agreement." Just for the record, Hobbes's italicized phrase does not cover all of the relationships giving rise to evidential signs. For example, a lightning flash is a sign of thunder (and vice versa) even though neither is the consequence of the other. In this case, both signifier and signified have a common cause. Noncausal correlations may also generate evidential signs.

5 The same goes for Arnauld and Locke. See Kretzmann 1967: 376–8; 1968; Hacking 1975; Hungerland & Vick 1981: Chapters 1–3; and Martinich 1981: 354. See also Horwich (1998a: 19–20), who argues that "means" has its evidential sense even in (2).

(5) S means to say "boulder." (*Intention*)
(6) S means what he says. (*Seriousness*)

Sentence (5) concerns what S intends to say, and hence may be true even though, owing to a slip of the tongue, or a speech impediment, S actually says "bowler." In the same sense, one may mean to win, or to be rich. Intending is also expressed when "means" is followed by "as." Thus "S meant his remark as a joke" means simply that S intended his remark to be a joke. Note that S may mean to say "boulder" even though he does not mean "boulder," as when (3) is true. Conversely, S may mean "boulder" without intending to say "boulder," as when S uses a synonym or a foreign language. And when S puts his fingers to his lips in a familiar gesture, he means "Be quiet!" without intending to say anything.

Finally, sentence (6) – or, more simply, "S means it" – claims that S's utterance is intentional, literal, and serious. Hence when S says "I moved some boulders," (6) may be false because he is exaggerating or joking, even when he means "I moved some boulders" rather than "I moved some cocaine." Sentence (6) would also be false if S said "I moved some boulders" by a slip of the tongue. Meaning it is compatible with lying, however.[6] Indeed, a witness would not be guilty of perjury if she did not mean it when she said that the defendant was with her all night. Before the prosecutor charges perjury, he must be sure that the witness meant exactly what she said. "Means" has senses even more distant from its linguistic sense in "My life has no meaning" and "His love means a lot to me."

The different senses of signification have parallels among implication, expression, indication, reference, signaling, and so on.

Implication

Evidential: The fact that S said "Mozart wrote over four hundred pieces of music" implies that S is alive.
Word: "Mozart wrote over four hundred pieces of music" implies "The natural logarithm of Mozart's last opus number exceeds 5.99."
Speaker: By saying "Mozart wrote over four hundred pieces of music," S implied that Mozart was a genius.

Indication

Evidential: The blinking light indicates that S is going to turn.
Word: N/A.
Speaker: By turning on his blinker, S indicated that he is going to turn.

6 I am indebted here to Mark Siebel and Christian Plunze.

Evidential implication and signaling appear to be identical to evidential meaning, while evidential indication is a weaker cousin.

In the process of defining speaker meaning, we will also define speaker implication and expression. All three depend on the speaker's intentions. They depend specifically, I shall argue, on what the speaker intends his actions to indicate about his beliefs and other mental states, in the evidential sense of indication. All three allow insincerity, cases where the speaker lacks the indicated mental state. What words mean, imply, or express depends not on any particular speaker's intentions, but rather on conventional usage and the rules of the language, which are determined by the intentions of prior speakers of the language. Evidential meaning, implication, and expression do not in general depend on intentions at all, and so exist outside the realm of intelligent, language-using creatures. The analyses I shall present, therefore, will show that all five senses of meaning that we have distinguished have intricate interconnections.

§2.2 COGITATIVE VERSUS COGNITIVE SPEAKER MEANING

Schiffer (1972: 2–3) showed that there are even two kinds of speaker meaning.

(7) By (the expression) e, S meant "p." (*Cogitative Speaker Meaning*)
(8) By (saying or doing) e, S meant that p. (*Cognitive Speaker Meaning*)

We will focus on what S meant on a particular occasion, not on what he used to mean in a frequentive sense (cf. §11.6). Forms (7) and (8) are not as obviously distinct, and often have the same truth value. Indeed, "S said 'p' and meant it" entails both (7) and (8) when e is the same sentence as "p." Nevertheless, one difference is that in (7) any meaningful expression can replace "p," whereas in (8) only declarative sentences can replace "p." Thus S can *mean* "a female fox," "Go to the store!" or "Is that a canary?" as well as "Today is Sunday." But among these, S can only mean *that* today is Sunday. "S meant that a female fox," for example, is ill formed and unintelligible.

Another difference – linguistically subtle and theoretically insignificant, yet real nonetheless – is that the parenthetical insertions in (7) and (8) cannot be switched without either impropriety or a change of meaning. The novelist meant "Nuclear bombs were dropped on New York City" *by the expression* "New York was nuked," not *by saying* "New York was nuked." On the other hand, S meant that Charlie is obese by saying "Charlie is

a sea elephant," not by the expression "Charlie is a sea elephant." The term "expression" here can be replaced by something more specific like "words" or "sentence," or by cognate terms like "signal," "sign," and "gesture." And "saying" can be replaced by "uttering," "writing," "declaring," "signing," and so on. Note also that only (8) can be reexpressed as "S meant that p *when he said* e."[7]

More significantly, the fact that (7) is *oratio recta* while (8) is *oratio obliqua* means that we have to adjust for indexicals. When Steve said "I cut myself," he meant *that he cut himself*, assuming that he was using English normally. He did not mean *that I cut myself*, since Steve cannot use the first person pronoun to talk about me. But by the sentence "I cut myself" Steve meant "I cut myself," *not* "He cut himself." Hence (7) and (8) will normally not both be true when exactly the same sentence with indexicals is substituted for "p" in both places.[8]

The most important point is that (7) and (8) have different truth conditions even when we restrict "p" to pronoun-free declarative sentences. Sentence (7) does not entail sentence (8), and (8) does not entail (7). This can be seen in the cases of *fiction* and *metaphor* described earlier, as well as in the following.

Irony. Suppose that S said "Rockefeller is in hock," meaning "Rockefeller is in debt" by that sentence rather than "Rockefeller is in white Rhine wine." As Schiffer (1972: 2–3) observed, S did not mean that Rockefeller is in debt if he was speaking ironically. Indeed, S meant that Rockefeller is in great financial shape.

The referential use of definite descriptions.[9] By saying "The man drinking the martini is drunk," S may have meant that Smith is drunk, even though he did not mean "Smith is drunk" by the words he used. As a consequence, S can be accused of having made a mistake of some sort if it turns out that while Smith is drunk, he is not drinking a martini. If S *had* meant "Smith is drunk" by his words, then he deviated from convention but did not make a mistake.

Parts of compound sentences. A speaker who says "If I won the lottery, I am a millionaire" would normally mean "I won the lottery" by the antecedent of the conditional he uttered. But he would normally not mean by uttering it that he won.

7 The locution *S meant "p" by saying e* (or *when he said e*) has some features of (7) but is used with the meaning of (8). *By e, S meant "p"* is ambiguous, and can mean either (7) or (8). For clarity, I will use *By . . . S meant "p"* only when I have the sense of (7) in mind.

8 This was observed by S. Davis (1994) in the case of word meaning.

9 See Donnellan 1966; Kripke 1977; Bertolet 1987; and §6.1 of this volume.

What is the difference between (7) and (8) when "p" is a declarative sentence (ignoring or adjusting for indexicals)? My suggestion is that (8) entails that S used e to *express the belief* that p, while (7) entails only that S used e to express the *thought* or *proposition* that p. Accordingly, I say that (8) expresses *cognitive* speaker meaning, while (7) expresses *cogitative* speaker meaning. Our novelist used "New York was nuked" to express the thought, not the belief, that nuclear bombs were dropped on New York. To the extent that an author is expressing beliefs in a literary work, it ceases to be a work of fiction.

The relevant notion of thought will be discussed at length in Part III. For now, observe that the word "thought" is at least doubly ambiguous. It suffers initially from the familiar *act-object ambiguity*. "S's thought that p" can refer either to what S thinks, or to S's act of thinking it. What S thinks, the content of his thought, is simply the proposition that p. In addition, "thought" qua act can mean either *believing* or *occurrent thinking*. In "S *thinks* that he will go to the party," "thinks" means "believes." "Think" has the other meaning – the one we shall intend – in "S is thinking the thought that he will go"; "S is thinking about going"; "S is thinking about the party"; "S's train of thought was interrupted"; "S is thinking pleasant thoughts," and so on. "S is thinking a thought" is equivalent to "The thought is *occurring* to S." When a thought occurs to us, we say that it *comes* or is *brought to mind*, or that it *crosses* or *enters our mind*.

Thought in this cogitative sense is cognitively neutral. That is, S may be thinking the thought that it is raining in Siam whether he believes or disbelieves that it is raining there. Conversely, S may believe that $2 + 2 = 4$ whether or not it is occurring to him at the moment that $2 + 2 = 4$. Indeed, we have all believed that $2 + 2 = 4$ since childhood, even though the thought has occurred to us only on selected occasions, as when balancing our checkbooks. When a subject is thinking a proposition that he believes, he is said to *occurrently believe* that proposition.

Whereas beliefs are normally ascribed using *oratio obliqua* alone, thoughts are just as often ascribed using *oratio recta*. "Steve is thinking the thought 'I cut myself' " (*oratio recta*) is equivalent to "Steve is thinking the thought that he cut himself" (*oratio obliqua*). Thus we can say that Steve meant "I cut myself" by "I cut myself" either because he used the sentence to express the thought "I cut myself" or, equivalently, because he used the sentence to express the thought that he cut himself.

Guessing presents a problem for the suggested analysis of cognitive speaker meaning. Suppose that Ed asks, "Exactly how many beans are in the jar?" Mike replies, "I don't know." When asked to "just guess," he

says "5,639." Mike's answer was obviously elliptical. If he meant that there are 5,639 beans in the jar,[10] then expression of belief is not necessary for cognitive speaker meaning. For while Mike may have used "5,639" to express *some degree of certainty* that there are 5,639 beans in the jar (it's not a guess if you're completely certain that it's wrong), he surely did not use it to express the *belief* that there are (it's also not just a guess if you believe that it's right).

It is plausible that by "5,639," Mike meant "There are 5,639 beans in the jar." But we have seen that a speaker often means something cogitatively without meaning it cognitively. So we cannot automatically infer that Mike meant *that* there are 5,639 beans in the jar. If he did, he could be accused of being foolish, of making claims for which he had insufficient evidence. But surely he wasn't. If Bill said to Mike, "You fool! How could you maintain that the jar contains 5,639 beans when you haven't counted them?" Mike could justly reply, "I never said it did." If Bill pressed further, saying "But you said '5,639' when Ed asked how many beans the jar contains," Mike could reply, "That's true. But Ed asked me to make a guess." Mike may have meant that *it is his guess* that there are 5,639 beans, or that there are *about* 5,639 beans. But these are things that he should believe, at least after making his guess.

Mark Lance has suggested that S can mean that p without expressing the belief that p when "playing devil's advocate." Consider the following dialogue.

Alan: Bush is simply a fool to go ahead with the Strategic Defense Initiative in violation of the ABM treaty with the Russians.

Bob: I agree with you, but let me play devil's advocate. Here is an argument that Bush is doing the right thing. Nuclear weapons, and especially nuclear technology, are widely available. The same goes for ballistic missiles. There is a real possibility, therefore, that someone other than the Russians will launch a rogue nuclear missile at us. Therefore, Bush is acting wisely.

Did Bob mean that Bush is acting wisely by anything that he said? I do not believe that we can say "Yes" without misinterpreting Bob, and ignoring the fact that he is *just* playing devil's advocate. It is clear that Bob meant "Bush is acting wisely" by the last words he uttered. But that is a statement about cogitative rather than cognitive speaker meaning. It is also clear that Bob *said that* Bush is acting wisely. But I think it would be a misinterpretation to claim that Bob *said and meant* that Bush is acting wisely. This is

10 Cf. Searle 1979: 12–3; Vlach 1981: 369; and Martinich 1984a: 126–7.

not like the case in which Bob graciously says "You are doing a fine job" to Bush at a party. In that case, Bob said and meant that Bush is doing a fine job, even though he may not have really meant what he said.

§2.3 MEANING, IMPLICATION, AND EXPRESSION

One final division: there is both an *inclusive* and an *exclusive* sense of cognitive speaker meaning. In the latter, meaning *excludes* implying. By "It is necessary that $2+2=4$" we imply, but do not mean, that $2+2$ equals 4. We mean rather than imply that $2+2$ necessarily equals 4. By "Bob's son is a thief," we mean rather than imply that he is. We may imply but not mean: that Bob has a son, that Bob's son is morally reprehensible, that we can back up our claim, and so on. Finally, a student who says "I have to study" in response to "Do you want to go to the movies?" implies, but does not mean, that he does not want to go. He means rather than implies that he has to study.[11]

Implying that p also entails expressing the belief that p. When a student says "I have to study," he expresses both the belief that he has to study and the belief that he does not want to go to the movies. Indeed, he expresses the latter belief *by* expressing the former. Let us say that S *directly* expresses the belief that p provided S expresses the belief that p but not by expressing another belief. S expresses a belief *indirectly* if he does so by expressing another belief. Then we use "It is necessary that $2+2=4$" to directly express the belief that $2+2$ necessarily equals 4, and to indirectly express the belief that $2+2$ equals 4. Whereas meaning in the exclusive sense requires direct expression, implication requires indirect expression. Indeed, meaning and implication are mutually exclusive and jointly exhaustive species of expression.[12]

11 "Indirect speech acts" (Austin 1962: 129; Searle 1969: 65; 1979: 30–57; Tormey 1971: 70; Holdcroft 1978: 61–3; Carr 1978a; Fasold 1990: 152–7; Chierchia & McConnell-Ginet 1990: 161) provide numerous examples of the contrast between meaning and implication, as do "presuppositions" (Bach & Harnish 1979: 155–72) and other "implicatures" (Grice 1975; Martinich 1984a: Chapter IV; Chierchia & McConnell-Ginet 1990: 187–203; Davis 1998), and the referential use of definite descriptions (McKinsey 1978: 177). Grice, however, opposes what a speaker "implicates" to what he has *said*, and so counts ironic meaning as an implicature. In the sense we are examining, someone who says, ironically, "That's wonderful!" means rather than implies that it is the very opposite of wonderful.

12 Conjunctions are a borderline case. Suppose S says "Bob cooked and Bill cleaned." Did S merely imply that Bill cleaned? Or did he mean that Bill cleaned? Did S express belief in both conjuncts by expressing belief in the conjunction? Or vice versa? I am inclined to answer the first and third questions "Yes" but see no way to definitely resolve the issue.

In its inclusive sense, meaning *includes* implying, and involves the expression of belief either directly or indirectly. It would be natural in many contexts to describe the student as having meant that he did not want to go when he said "I have to study." When the subtle direct-indirect distinction is unimportant, it would be impertinent to insist that the student implied, but did not mean, that he did not want to go. The term "means" is thus like "animal," which has a general sense in which it applies to humans, and a more specific sense in which it does not. "Implies," by contrast, always connotes indirectness of expression.

Having noted the inclusive sense of "means," we will henceforth focus on the exclusive sense. We will accordingly formulate our definitions as follows:

2.1 **Definition:** *S means that p iff S directly expresses the belief that p.*

2.2 **Definition:** *S implies that p iff S indirectly expresses the belief that p.*

These definitions jointly imply:

2.3 **Theorem:** *S expresses the belief that p iff S either means or implies that p.*

It also follows as a theorem that S means that p *by saying e* iff S expresses the belief that p *in that way*. We earlier noted that "S means that p by saying e" is often abbreviated by dropping the "saying."

Starkly different act trees, with resultant differences in what is meant as opposed to what is implied, are represented by standard uses of the following:

(9) It will rain.
(10) It will rain, I believe.
(11) I believe that it will rain.

All three sentences can be used to express both the belief that it will rain (R) and the belief that one believes that it will rain (B). But the user of (9) typically expresses B by expressing R. Hence he means that it will rain, and implies that he believes that it will rain. The typical user of (11), by contrast, expresses R by expressing B, and accordingly means that he believes that it will rain, and implies that it will rain. The user of (10) does not express R by expressing B, nor does he express B by expressing R. Hence he means both that it will rain, and that he believes that it will

It should hardly be surprising that the terms we seek to define are somewhat vague. Nearly all terms are vague to some extent.

rain. When the speaker uses (9) to express R but not B, then he neither means nor implies that he believes that it will rain.[13]

A natural conjecture at this point is that what S meant in the inclusive sense is comprised of what S *said or implied*, and that what S meant in the exclusive sense coincides with what S *said*.[14] Indeed, in examples (9) through (11), what the typical S says coincides with the belief that S directly expresses. In using (9), we imply but do not say that we believe that it will rain; what we say is that it will rain. In using (11), we imply but do not say that it will rain; what we say is that we believe that it will rain. There is at least one case, however, in which speakers mean something in both senses that they neither say nor imply, and say something without meaning or implying it. That is the case in which the speaker says something by mistake, through either a slip of the tongue or linguistic ignorance. If Bill says "Dr. Johnson is an etymologist," either because he thinks that "etymologist" means "entomologist" (insect scientist), or because he intended to say "entomologist" but the wrong word came out, then Bill said that Dr. Johnson is an etymologist (word scientist) even though he neither meant nor implied this. He meant that Dr. Johnson is an entomologist even though he neither said nor implied that. What a speaker says is more closely connected than what he either meant or implied to what his words mean. What we cannot do by mistake, I believe, is to *assert* (*state, affirm, propose*, or *declare*) that something is the case. If a scholar said or wrote "p" through either a slip of the tongue or linguistic error, then we cannot count that scholar as among those who have asserted that p. To assert that p, we must both say and mean that p.

§2.4 COGITATIVE SPEAKER MEANING (EXCLUSIVE)

In the cogitative form, "mean" also has both an *inclusive* and an *exclusive* sense,[15] although "implied" has no cognate cogitative sense. Whether "By

13 There are additional contrasts, of course. For example, (9) is typically used to express certainty about rain, while (10) is typically used to express uncertainty; (11) may be used without implying certainty or uncertainty about rain.

14 Cf. Neale's (1992: 520) tentative exegesis of Grice's notion of the "total signification" of an utterance. In at least one place, Grice does explicitly state that S's saying that p entails S's meaning that p (1989: 87). He suggests the view in 1968: 227–9 and 1989: 120–1 without endorsing it, and omits even the suggestion from the original version of "Utterer's Meaning and Intention" (1969). See also Neale 1992: 554 and Saul 2001. I am indebted to Jennifer Saul for bringing these passages in Grice to my attention.

15 I failed to recognize the inclusive sense in previous publications (Davis 1992a, 1992b), even though my critique of Schiffer's analysis (1992b: 83) relied on it. As a result, I was bothered

(the expression) e, S meant 'p' " has the exclusive or the inclusive sense depends on whether "the expression" (or a cognate term) is present or absent. In the exclusive sense, we can say that by *the sentence* "Rockefeller is in hock," Steve meant "Rockefeller is in debt," not "Rockefeller is in great financial shape." In the inclusive sense, we can say that S meant "Rockefeller is in debt" by "Rockefeller is in hock," and thereby meant "Rockefeller is in great financial shape." Pronouns and ellipses provide further examples. In the inclusive sense, Tanya might well mean "Edberg won the U.S. Open" by "He won." But Tanya would not mean "Edberg won the U.S. Open" by *the sentence* "He won," unless she were speaking in a code of some sort. What speakers normally mean by the sentence "He won" is just "He won." Whereas the exclusive sense is more closely tied to word meaning, the inclusive sense is closer to speaker reference (§6.1). Tanya would not ordinarily mean "Edberg" by the pronoun "he," but she would have meant Edberg if that was who she was referring to.

As in the cognitive case, the difference between the inclusive and exclusive senses depends on whether the expression may be indirect or not. We will focus on the exclusive sense. Let "ρ" stand for a pronoun-free declarative sentence expressing a proposition. For this range of substitution instances, the definition of cogitative meaning goes like Definition 2.1, with thought replacing belief. The thought that ρ can also be referred to as the *idea* that ρ.

2.4 **Theorem:** *By (the expression) e, S means "ρ" iff S directly expresses the idea or thought that ρ by producing e.*

Theorem 2.4 follows from Definitions 2.5 and 7.7. When Steve uses "Rockefeller is in hock" ironically, he expresses both the thought that Rockefeller is in debt *and* the thought that Rockefeller is in great financial shape. But he expresses the latter thought *by* expressing the former. So by the sentence "Rockefeller is in hock," Steve meant "Rockefeller is in debt," not "Rockefeller is in great financial shape." By contrast, suppose that Tanya is using a code in which "hock" means "great financial shape," so that by the sentence "Rockefeller is in hock" she means "Rockefeller is in great financial shape." Then Tanya is not using irony. This shows that irony is not simply the use of a sentence to mean the opposite of what the sentence means conventionally. Rather, irony is the use of a sentence to express one proposition directly in order to express a contrary proposition

by conflicting intuitions, and overstated the limitations of Schiffer's analysis (cf. §4.5 of this volume).

indirectly, so that cognitive speaker meaning is the opposite of cogitative (in the exclusive senses). The belief expressed is the opposite of what the speaker means by the sentence. Thus if Tanya were to speak ironically in her code, she could use "Rockefeller is in hock" to mean that Rockefeller is in debt!

The difference between "dead" and "live" metaphors turns on whether the speaker's thought is expressed indirectly. When S said "Charlie is a sea elephant," his metaphor was live. S expressed the thought that Charlie is obese by expressing the thought that he is a certain sort of especially blubbery seal. Hence S meant "Charlie is a certain sort of seal" rather than "Charlie is obese" by the sentence "Charlie is a sea elephant." By contrast, when I say "There is a fork in the road," I am using a dead metaphor. I express the thought that there is a bifurcation in the road, but not by expressing the thought that there is a pronged utensil in the road, or the thought that the road is such a utensil. The idea of such a utensil generally does not even cross my mind. Hence I mean "There is a bifurcation in the road" rather than "There is a pronged utensil in the road" by the words I uttered.

Even in the exclusive sense, cognitive speaker meaning allows more indirection than cogitative speaker meaning. As a result, direct expression must be defined more strictly for thought than for belief if Definition 2.1 and Theorem 2.4 are to give the correct results. A *belief* is expressed directly, we will say, if it is not expressed by expressing any other *belief*. This allows the belief to be expressed by expressing a mental state other than belief, such as thought. In the case of irony, as we have seen, Steve expressed the belief that Rockefeller is in great financial shape by expressing the thought that he is in debt. Nevertheless, Steve meant, rather than implied, that Rockefeller is in great shape. Or suppose that S answers "Ow!" when asked "How do you feel?" Then S simulated an expression of pain, and thereby expressed the belief that he is in pain. While there is an element of indirection here, S meant rather than implied that he is in pain. As these cases illustrate, the exclusive sense of cognitive speaker meaning, which is opposed to implying, excludes only the expression of a belief by the expression of another belief. The inclusive sense, of course, allows even that.

A *thought* is expressed directly, by contrast, if it is not expressed by expressing any other *mental state*. A thought expressed by means of another thought is therefore expressed indirectly, as was the thought of Rockefeller's being in great financial shape in the irony case. Moreover, a thought expressed by expressing an entirely different mental state,

such as pain, is also expressed indirectly. When S said "Ow!" he expressed pain and thereby expressed the thought, as well as the belief, that he is in pain. Then by our stipulation, S expressed the belief directly, and the thought indirectly. Hence Theorem 2.4 correctly rules that S did not mean "I am in pain" by the one-word sentence "Ow!" And Definition 2.1 correctly rules that S did mean that he is in pain by saying it.

In a well-known example, Ziff (1967) imagined an irritable academic, George, who is compelled on induction into the army to take a moronic test designed to establish sanity. In response to "What would you say if you were asked to identify yourself?" George replies "Ugh ugh blugh blugh." Ziff observed that while George may have meant something by uttering this, he did not mean anything by "Ugh ugh blugh blugh." I would say that George expressed irritation by uttering "Ugh ugh blugh blugh" and thereby expressed both the thought and the belief that the test is stupid. Since George directly expressed that belief, Definition 2.1 rules that George meant that the test was stupid. Since George did not directly express the thought, Theorem 2.4 rules that George did not mean "The test is stupid" by the words "Ugh ugh blugh blugh." As a contrasting case, let us imagine that Gail is slightly deranged, and actually believes that "Ugh ugh blugh blugh" means "This test is stupid." When asked the same question, Gail gives the same answer. Then while Gail may have expressed irritation as well as the thought that the exam is stupid, she did not express the latter by expressing the former. Indeed, the reverse holds: she expressed irritation by expressing the thought that the exam is stupid. Since Gail, unlike George, directly expressed this thought, Gail meant "The exam is stupid" by the words "Ugh ugh blugh blugh."

We have focused on instances in which "p" is replaced by a declarative sentence. While the cognitive form (8) has no other instances, "p" can be replaced by any grammatical and meaningful unit in the cogitative form (7). When "p" is an interrogative, imperative, or performative sentence, we can still say that S means "p" provided that S directly expressed a thought, the thought "p." But the thought in this case will be *nonpropositional*. Thus S may directly express the thought "What time is it?" when he says "¿Que hora est?" To include the case in which "p" is not a complete sentence, Theorem 2.4 can be generalized to say that S meant "p" provided that S directly expressed the *idea* (or *concept*) "p," where ideas (concepts) are defined as *thoughts or parts of thoughts*. Thus S may have expressed the idea of a female fox by uttering the word "vixen."

The thought that no cats are vixens is a complex entity, containing the idea of a cat as well as that of a vixen. The idea of a vixen, moreover, is a component of many different thoughts, such as the thought that no dogs are vixens, the thought that all vixens are mammals, and so on. A thought occurs to someone only if all of its component ideas do.

We will use the terms "idea" and "concept" interchangeably. The nature of thoughts and ideas will be clarified in Chapters 12–20, and ideational theories of meaning will be defended against common objections in Chapters 21–23. Briefly, ideas are event-types of a certain sort, specifically, thoughts or parts of thoughts. Hence ideas are mental representations. The change in a person who is thinking "All cats are mammals" one second and "All dogs are mammals" the next is a change in the ideas that are occurring to him. Both of the elementary ideas that changed are occurring to someone who is thinking the thought that all cats are dogs. The same goes for the thought that all dogs are cats, which is a differently structured event-type in which the same component ideas stand in a different relation to each other. Some thoughts are propositions, and are the objects of belief, desire, intention, and the other propositional attitudes.

Ideas in the sense of thought-parts are to be sharply distinguished from two other sorts of mental representations, namely *images* and *conceptions* (Chapter 19). Images are complexes of sensations. Conceptions are belief systems. The images and conceptions associated with an expression are important determinants of its usage. Nevertheless, meaning is not determined by either, and does not consist in their expression. First of all, words are not used to express images. We do not use the word "cat" to express the image of a cat, or an image of the word, although we do use words to refer to and describe images. Words are used to express conceptions, but we need to use expressions with the complexity of a theory or an essay, not individual words. More importantly, meanings do not stand in a one-to-one correspondence with either images or conceptions. For example, our conception of a three-sided polygon is identical to our conception of a three-angled polygon, and any image we form of the former is an image of the latter. But we use "three-sided polygon" and "three-angled polygon" to express different thought-parts. Hence what we mean by "three-sided polygon" is different from what we mean by "three-angled polygon." Conversely, we have meant the same thing by "Jupiter" for centuries. We have used it to express the same thought-part. But our conception and image of Jupiter has changed dramatically. We used to believe that Jupiter revolved around the Earth, and was an unchanging sphere. Now we know that it revolves around the Sun, and has visible and colorful storms raging

Table 2.1. *The placeholder "μ"*

means μ	the idea μ	μ
means "female fox"	the idea "female fox"	"female fox"
means *female fox*	the idea *female fox*	*female fox*
means female fox	the idea female fox	female fox
means "Dogs bark"	the idea "Dogs bark"	"Dogs bark"
means *Dogs bark*	the idea *Dogs bark*	*Dogs bark*
means Dogs bark	the idea Dogs bark	Dogs bark
means "Get money!"	the idea "Get money!"	"Get money!"
means *Get money!*	the idea *Get money!*	*Get money!*
means Get money!	the idea Get money!	Get money!

on its surface. Finally, we definitely mean something by "if" and "and," which we use to express certain parts of thoughts. But such words are not associated with either images or conceptions.

We observed earlier that when it expresses cogitative rather than cognitive speaker meaning, "means" can be followed by individual words and phrases as well as by full sentences. And it can be followed by interrogative or imperative sentences as well as by declarative sentences. Since "p" is customarily used as a placeholder for declarative sentences only, it would be confusing to use it more generally. We will therefore introduce the Greek letter "μ" as a placeholder for the full range of expressions that can follow "means" when it expresses cogitative speaker meaning. The same range of expressions can follow "idea" and "concept," as well as "means" when it expresses word meaning. Thus "vixen" means "female fox," and S can mean "female fox" by the word "vixen." The word means "female fox" because it expresses the idea "female fox," and S means "female fox" by it when he expresses the idea "female fox" (see Table 2.1). As Table 2.1 indicates, the expressions following "means" and "idea" can be either quoted or unquoted, and in either italic or roman font; the meaning is exactly the same. Instances of "μ" could also be both italicized and quoted, but that is uncommon. These options are not available after "that," unless quotes are used as scare quotes and italics is used for the purpose of emphasis, functions they do not have in Table 2.1.

Quotation marks after "means" and "idea" are not used to form the names of the words quoted. *"Vixen" means "female fox"* does not say that

"vixen" means *the phrase "female fox."* On the left of "means," quotation is used to form the names of expressions. *"Vixen" means "female fox"* does say that *the word "vixen"* means "female fox." A full explanation of the notion will not be possible until later (§7.6), but we will say that the expressions following "means" and "idea" are used *ideo-reflexively*: they are used both to *express* an idea and to *refer to* it. Thus *By "vixen," Steve meant "female fox"* entails that Steve used "vixen" to *express the idea "female fox."* Like "vixen," most words are ambiguous, and therefore express more than one idea in English. On any occasion in which "the idea 'vixen' " is used, however, the word "vixen" is used with a particular meaning. "The idea 'vixen' " is used to refer to the idea that "vixen" expresses on that occasion. Thus when we are talking about animals, "the idea 'vixen' " designates the idea of a female fox. When we are talking about people, the same phrase refers to the idea of a shrewish woman. In general: *on any occasion of use, "the idea μ" refers to the idea expressed by "μ" on that occasion, if there is one.* This rule holds because expressions of the form "the idea μ" are indexical descriptions used deictically.[16] A substitution instance of "the idea μ" has the referent it has because the speaker is aware of the idea expressed by "μ"on that occasion. The term "ideo-reflexive" stresses the principal similarities and differences between this form of indexicality and the "token-reflexive" form. The distinction between *expressing* and *referring* is crucial: expressing will be defined for speakers in Chapter 3, referring in Chapter 6. Both notions will be defined for words in Chapter 7.

In order to define cogitative speaker meaning in terms of expression, we need to index "μ" with variables and constants standing for ideas or other mental states that can be expressed. Let "μ_i" stand for instances of "μ" that express i on that occasion. If f is the idea of a female fox, then "means 'vixen' " is a substitution instance of "means μ_f" on a given occasion only if "vixen" expresses the idea of a female fox on that occasion. The same goes for "means *vixen*" and "means vixen." "Means 'female fox' " is also a substitution instance of "means μ_f," assuming that "female fox" has its salient English sense. If s is the idea of a shrewish woman, then "means 'vixen' " is a substitution instance of "means μ_s" on a given occasion only if "vixen" expresses the idea of a shrewish woman on that occasion. "Means 'female fox' " could not be a substitution instance of "means μ_s" unless we were using a code of some sort. We can now use the following schema to define cogitative speaker meaning.

16 See my *Nondescriptive Meaning and Reference.*

2.5 **Definition:** *By (the expression) e, S means μ_i iff S directly expresses i by producing e.*[17]

Thus Steve means "debt" by "hock" iff Steve directly expressed the idea "debt" by producing "hock," and so on for all replacements for the placeholder "μ_i." "Directly" is defined for "idea" as for thought: an idea is expressed directly provided it is not expressed by expressing any other mental state. It will be noticed that the meaning of the subscript in Definition 2.5 was given in terms of what a word expresses on a given occasion of use. No circularity will result from this, since what a word expresses on an occasion, like speaker meaning and word meaning, will eventually be defined in terms of speaker expression, not speaker meaning.[18]

The placeholder "ρ" was restricted to pronoun-free sentences in Theorem 2.4 because pronouns behave differently in *oratio recta* and *oratio obliqua*. Jane will typically use "I am a woman" to express the thought "I am a woman," but not the thought that I am a woman. Since she means "I am a woman" without thinking the thought that I am a woman, Theorem 2.4 would be false if "I am a woman" were an allowable substitution for "ρ." Conversely, Jane did express the thought that *she herself* is a woman, but did not mean "She herself is a woman" by the sentence "I am a woman."[19]

We are using the letter "e" as the variable ranging over potential signifiers because linguistic expressions are its typical value. But "e" should be interpreted without restriction as a general individual variable. It ranges over any action or object that an agent might or might not use to express an idea. Thus "e expresses i" may be true when e is a gesture,

17 As explained more fully in §6.4, "μ_i" is a substitutional variable in Definition 2.5, whereas "e" and "i" are individual variables. All three are bound by implicit universal quantifiers of the appropriate type: substitutional for "μ_i," objectual for "e" and "i." A substitution instance of a form containing "μ_i" is obtained by uniformly replacing "μ_i" with any of its permissible substituends: any English expression expressing i, or a constant stipulated to express i.

18 A definition for the inclusive sense of cogitative meaning will require more than simply dropping "directly" from Theorem 2.4. In a case discussed in *Nondescriptive Meaning and Reference*, the president said, "We must remember that always we have a duty to protect the lives of American citizens abroad," alluding to the invasion of Grenada. Since he made an oblique reference to Grenada, he expressed the idea of Grenada. But there seems to be no sense in which the president *meant* "Grenada" by "We must. . . ." The inclusive sense of cogitative meaning allows most forms of indirection but not all.

19 We might suspect that "ρ" needs to be completely free of indexicality. But "S is thinking the thought 'Tanya is a foreigner'" and "S is thinking the thought that Tanya is a foreigner" appear to be equivalent despite the indexicality of "foreigner" and "is."

hand signal, facial expression, noise, mark, or sign, as well as when it is a written or spoken word. And "e does not express i" will be true when e is an asteroid or quark. (Of course, we can use a definite description of the form "the expression e" only when e is an expression, but that is not because we have restricted the range of "e" to expressions.) We will similarly use the word "produce" quite generally, so that we can be said to produce any gesture, hand signal, facial expression, noise, or mark by which we might mean something, as well as written or spoken words. We can produce e by saying, writing, performing, presenting, or making e.

There is a semantic difference between meaning and expression that is not implied by our definitions.

> One can ask, "Why did you refer to him?", but not "Why did you mean him?" One can say, "Don't refer to him!", but not "Don't mean him!". . . . The verb "to mean" has noncontinuous present tense forms, for example, "I mean you," but it lacks the present progressive tense form, "I am meaning you." The verb "to refer" has a present progressive form, "I am referring to you," as well as a noncontinuous present form, "I refer to Adlai Stevenson." What these grammatical considerations show is that referring to someone is an action; meaning someone is not an action. (Linsky 1963: 78)[20]

Similar considerations show that expressing as well as implying are like referring in being actions. The fact that expressing but not meaning is an action does not falsify either Definition 2.1 or Definition 2.5, since they were stated as biconditionals. The conditions stated are neither too broad nor too narrow. The fact that meaning is not an action does show, however, that the definientia are not *synonymous* with the definienda. Hence the "iff" cannot be strengthened to "means." Our definitions do show us where to look for semantic definitions, however. According to the definition to be presented here, to express an idea is to *act in a certain way with a certain intention*. It also follows that one expresses an idea only if one *has a certain intention that motivates one to act in a certain way*. It is plausible that this latter condition, when filled out, specifies the meaning of "S means_____." On this analysis, meaning something is a species of intention, not a species of action. The fact that "means" is not an action verb thus has little importance for any of the issues that we will be discussing. Accordingly, we shall not make much of it.

20 See also Stampe 1968: 139–43, 162–3.

In the vast majority of cases, a speaker means something by an expression because he uses it to express an idea, and words mean something because they express ideas. But there are exceptions. Consider the standard use of an *interjection*, such as "Ouch!"[21] There is a marked semantic difference between "Ouch!" and "I am in pain!" even though both are used to express pain. If S expresses pain by saying "I am in pain," he does so *by* expressing an idea, namely, the idea that he is in pain. But when S says "Ouch!" in the typical manner, he appears to be expressing pain directly. He is normally not expressing the idea that he is in pain. There is, moreover, no other idea that he has expressed directly. It is not true that S meant "Ouch!" iff S expressed the idea "Ouch!" For S did mean "Ouch!" even though there is no such idea. Or suppose that S is having difficulty speaking for some reason. T says "Jimmy won the match." S makes an inarticulate sound. We might wonder whether S meant anything by that, and if so, what. Our best guess might be that he meant "Oh!" or perhaps "Wow!" If he meant "Oh!," he expressed surprise. If he meant "Wow!," he expressed amazement. Such mental states may involve ideas, but they are not themselves ideas. There is no such thing as the idea "Oh!" or "Wow!" I will say that interjections have *nonideational meaning*.

Consider next *pejorative terms* such as "wop," "Polack," "nigger," and the like. These are used to express both an idea and an *attitude*. Thus "nigger" is used to express a complex mental state containing the idea of a Negro and a negative attitude toward such people. It is plausible that the difference between meaning "nigger" and meaning "Negro" is not due to the expression of different ideas, but to the expression of an attitude in the first case that is not expressed in the second.[22]

Definition 2.5 has been formulated to allow cases of nonideational speaker meaning. We selected the letter "i" as our variable because, in the typical case, what is expressed is an idea. But we have not stipulated that the range of "i" is restricted to ideas. Definition 2.5 can perfectly well hold when i is a mental state other than an idea. Thus it might well be true that a Frenchman meant "Ouch!" by "Aïe!" because he used "Aïe!" to express pain or some other nonideational state.[23] This does not falsify

21 Cf. Meinong 1910: 26.

22 I will argue in *Nondescriptive Meaning and Reference* that the difference between "Negro" and "nigger" actually is ideational.

23 See the discussion of interjections in *Nondescriptive Meaning and Reference.*

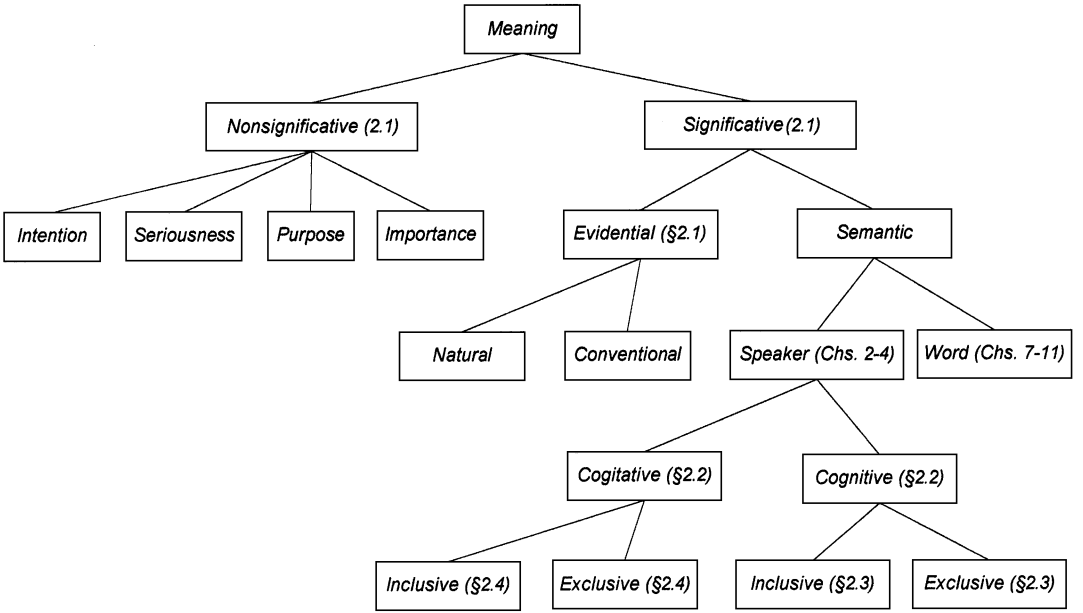


Figure 2.1. The senses of meaning

Definition 2.5, because "Ouch!" has the meaning it has in virtue of what it is used to express.

Consider next "Hello!" When used as a greeting, it expresses an emotion or attitude, something like "friendly regard."[24] But it is also used to answer a telephone. Answering a telephone is an overt action, not a mental state. It is something that is done, not expressed. So it might be thought that Definition 2.5 needs to be generalized even further. Indeed, we could say that *S meant* μ *iff S directly performed action D, where D is the action "*μ*" is used to perform*. In the typical case, D would be the act of expressing an idea or other mental state, but this formulation would allow purely performative meaning.[25] However, a case can be made that the use of "Hello!" to answer a phone is covered by Definition 2.5. For it seems that when one says "Hello!" upon picking up the phone, one is answering the phone *by* expressing friendly regard for the caller. Hence the action performed directly is the expression of a mental state. To see that this is the case, reflect on the fact that there are other ways to answer a telephone. One might say anything from "Bug off!" to "I love you!," not to mention "Thank you for using AT&T. How may I help you?" In these cases, one answers the phone by expressing rather different ideas.

I will suggest in §14.4 that punctuation marks such as commas and periods, when semantically significant rather than purely phonological, also have nonideational meaning, expressing purely structural features of thoughts.

We will concentrate on ideational meaning for two reasons. First, interjections, emotive terms, and punctuation marks represent an unusual and somewhat marginal case. Second, this book is motivated as much by a desire to understand thought and ideation as by the desire to understand meaning.

§2.6 THE SENSES OF MEANING

A tree representing all of the different senses of the word "meaning" is presented in Figure 2.1, along with the places where they are characterized in the text. Competing definitions of speaker meaning will be discussed in Chapter 4, after expression has been defined. Word meaning will be defined in Part II.

24 *Webster's Encyclopedic Unabridged Dictionary* (New York: Portland House, 1989).

25 Cf. Robins et al. (1997), who suggest that the purpose of language is to "cooperate" socially, rather than to express thought.

3

Expression

Having defined what a speaker means in terms of what the speaker expresses, we will now define speaker expression in terms of intention and evidential meaning. The basic idea is that to express a thought or other mental state is to provide an indication that it is occurring to us. We will distinguish speaker expression from closely related concepts, particularly that of evidential expression. Competing analyses of speaker expression will be discussed in Chapter 4. Word expression will be defined in Part II.

§3.1 SPEAKER, WORD, AND EVIDENTIAL SENSES

As we noted in Chapter 2, expression resembles meaning in having speaker, word, and evidential senses, related in roughly the same ways.

(1) The look on S's face expresses fear. (*Evidential Expression*)
(2) The word "fear" expresses the idea of fear. (*Word Expression*)
(3) By saying "I'm afraid," S is expressing fear. (*Speaker Expression*)

The first and most obvious difference concerns the subject to which the predicate "express" applies. In the speaker sense, the term applies to a person or other animate object, and does so in virtue of the person's saying or doing something. In the other two senses, the term "express" applies to inanimate objects like words or facial expressions, and not in virtue of their doing or saying anything themselves. Second, "x expressed fear" entails "x expressed *himself*" in the speaker sense, but words and facial expressions cannot express themselves. Third, "x expresses fear" entails "x is an expression" only in the nonspeaker senses. The look on S's face is an expression, and the word "fear" is another sort of expression. But S himself is not an expression in any sense, although his utterance

of "I'm afraid" may be an expression of fear in the evidential sense. The further distinction between word and evidential expression is marked by "of" and "for." An utterance or a fearful look is an expression *of* fear, whereas the word "fear" is an expression *for* fear.

A theoretically important difference concerns intentionality. To express fear in the speaker sense is to do something intentionally. Looks and words, on the other hand, cannot act intentionally. More significantly, facial expressions of emotion are typically involuntary manifestations of the emotion. And what the words in any language express is not dependent on the intentions of any particular speaker.

There are many connections between what words express and what speakers use the words to express. Thus the word "fear" expresses the idea of fear because speakers of English have conventionally used the word to express that idea. And I may use the word "fear" to express the idea of fear because that is what the word expresses. But speaker expression and word expression often differ radically. By a suitable stipulation, Steve could use the word "fear" to express the idea of happiness. And many English sentences express thoughts even though no one has ever actually used the sentences to express those thoughts. Because individual usage can differ from conventional usage, languages change over time as new uses catch on and become conventional.

Evidential expression does not depend essentially on either convention or intention. An evidential expression may thus occur as an unlearned response in animals lacking the ability to express themselves. The expression on a subject's face expresses fear because of the way it is caused by the fear. The expression may be involuntary and highly unusual. Evidential expression does not *exclude* conventionality or intentionality, either. For example, a woman's exclaiming "Yahoo!" may be an evidential expression of joy even though her exclamation is both intentional and conventional. In such cases, we have both speaker and evidential expression: the woman expressed joy, and her exclamation was an expression of joy.[1]

§3.2 INDICATION

Let us focus initially on what it is for a speaker to express a belief. We may begin with the fact that S can express the belief that p only by performing

1 The facts summarized in the previous paragraphs show that one sense of "express" cannot be described as narrower than another. Contrast Benson 1967: 340 and Koch 1983: 179. I define evidential expression in Davis 1988a.

some other action A. Typically, A is the act of using an expression that itself expresses the belief that p. More generally, A is an *indication* that S believes that p. Thus S may express the belief that it will rain by saying "It will rain" in a normal tone of voice, by saying "It's not going to rain" in an ironic tone, by nodding his head in response to "Will it rain?" and so on. Given the causal as well as statistical connection between thought and utterance, S's *saying* "It will rain" (or performing any of the other actions) is an indication that he *believes* that it will rain. That is, S's utterance is an *evidential sign* of his belief.[2] This is not to confuse evidential and symbolic meaning. The *sentence* "It will rain" means *symbolically* that *it will rain*. The fact that S *used* the sentence means or indicates *evidentially* that *S believes* that *it will rain*. This suggests that we may be able to define speaker expression, and thereby speaker meaning, in terms of evidential meaning or indication.

Evidential meaning is a bit too strong. For A means or implies B in the evidential sense only if B occurs or is true. "His temperature of 102 means that he is sick, but he is not sick" is a self-contradiction. So if expressing a belief entailed doing something that means evidentially that one has the belief, then lying or joking would be impossible. For these actions involve expressing beliefs that one does not hold. The classic philanderer who whispers "I love you" means that he loves you even though he doesn't "mean it." For similar reasons, expressing a belief does not entail *manifesting* it. An action manifests a belief only if the belief played a role in causing the action, which requires having that belief. *Making manifest* is too strong for a slightly different reason. Something is made manifest only if it is made obvious (evident, clear). But it is often far from obvious that a speaker believes what he says. A defendant who proclaims "I am innocent" has without question expressed the belief that he is innocent. Yet no one in the courtroom may be in a position to determine whether the defendant believes that he is innocent or not.[3]

A may be an *indication* of B, however, even though B is nonexistent or improbable. Thus the leading economic indicators may all indicate a

2 In addition to the classical sources cited in the Introduction, see Husserl 1900: 276–8, 296–7; Armstrong 1971; Stampe 1979; Hungerland & Vick 1981: 30–50, 127–48; Grice 1982; Blackburn 1984: 123–4; D. W. Smith 1982a: 195; Horwich 1998a: 19–20; Berg 1999: 80.

3 Contrast Sellars 1969: 102–3; Peacocke 1986: 115; and the condition attributed to Searle by Harnish 1976: 349. The idea that utterances "make manifest" the speaker's intentions plays a key role in the theory developed by Sperber and Wilson (1986a; 1987), who use the term without the connotation of obviousness. On their definition (1986a: 39), however, the term is much too weak to define meaning: an act will typically make manifest, in their sense, both that S believes p and that he does not. See also Evans 1982: 311.

booming economy even though the economy collapses. And a symptom might be an indication of disease even though the probability of disease given the symptom is less than 50 percent. Indications of an event are evidence that it has occurred, but the evidence need not be conclusive, and may not even be sufficient to make the event probable. Only when an indicator is perfectly reliable does it mean or imply what it indicates.[4] A liar wants his audience to think that he has a certain belief. So he produces an evidential sign that he has it, from which he hopes his audience will infer erroneously that he does.

Note carefully that "S expressed *the* belief that p" must be distinguished from "S expressed *his* belief that p." The latter implies that S in fact has the belief, and excludes insincerity; the former does not.[5] Similarly, "S expressed *his love* for Mary" implies that S really does love Mary; "S expressed *love* for Mary" does not. In general, "S expressed his ϕ" entails both "S expressed ϕ" and "S has ϕ," because singular terms of the form "S's ϕ" presuppose that S does have ϕ.

In the case of insincere expression, the action A in virtue of which S expresses a mental state cannot be an evidential expression of that state. For example, when S insincerely expresses love for Mary, his utterance cannot be described as a natural expression of love. For an evidential expression of ϕ implies that S does have ϕ. "His demeanor expressed fear but he was not afraid" is contradictory.

The expressive action must, I believe, be *publicly observable*. Suppose Steve says "It is going to rain" *out loud*, while Sam says the same thing *to himself*, not just "under his breath" but "in his head." Both actions may be an indication that the speaker believes it is going to rain. But only Steve has *expressed* his belief. Indeed, we say of people who are completely paralyzed or immobilized that they cannot express themselves at all. They may be able to think and feel, and to engage in inner speech. But they cannot communicate, signal, gesture, or perform any expressive actions. Only if people developed the ability to "read minds" could we express ourselves through inner speech.

What holds for expression in this regard holds for meaning, too, confirming Definition 2.1. While Steve has done something by which he means that it will rain, Sam has not. "By saying e, S meant that p" implies that S said e out loud. If we are given that S said to himself "It's in the

4 Cf. Bach & Harnish 1979: 58 and Recanati 1986: 221. I shall use "indicate" interchangeably with "is an indication of," although the former may be stronger in ordinary usage.

5 Contrast Tormey 1971: 39, 57–8, 68.

bank," it would be natural to ask what S was thinking, not what he meant. The words S said to himself have meaning, of course, and S may generally mean something by them. But neither fact would imply that S meant anything by saying them to himself on that occasion.[6]

While expressive actions must be observable, they need not be observed. And an action may indicate a condition even though no one realizes it, in which case the action does not indicate that condition *to* anyone.[7] Geologists did not realize that a certain type of mountain range was an indication of continental drift before plate tectonics was accepted in the twentieth century. But the indications were there all along. "A indicates B" says roughly that there is a causal or statistical relation between A and B in virtue of which A would give a suitably placed observer a reason to expect B. That condition can obtain even though no one is suitably placed. The fact that S's action may be an indication that he has a certain belief even though the action does not indicate that belief to anyone is important in connection with private diaries and secret codes (see §4.1).

§3.3 INTENTION

Producing an observable indication is not sufficient for expression, however. Normally, we would not spend money on insurance unless we believed driving to be risky. Consequently, the buying of car insurance is an evidential sign, and manifestation, of the belief that driving is risky. But we do not normally express that belief, or mean anything, by purchasing insurance. One thing missing is *intention*. When buying insurance, we perform an action that indicates a certain belief, but not on purpose. The action may be purposive, but the fact that it indicates the belief is not. By contrast, one who predicts rain does purposely perform an action indicating a belief that it will rain.

Some animal communication illustrates the same point. According to Denkel (1983), if a vervet monkey sees an eagle swooping down, the monkey barks in alarm. All monkeys in the group then drop to the ground for protection in the undergrowth. If a leopard is spotted, a different bark is produced, and the monkeys seek safety higher in the trees. The monkey's barking clearly indicates that an eagle (or leopard) is approaching, and that

6 Cf. Husserl 1900: 279; Harman 1968: 590–6; and Sellars 1969: 104–5. The intention condition discussed in §3.3 also fails in the case of inner speech. But it could conceivably be satisfied, without the subject expressing himself or meaning anything.

7 Contrast Husserl 1900: 270.

the monkey is aware of its approach. But we would hesitate to say that the monkey is expressing the belief that an eagle is approaching. For that requires quite a bit of self-reflection and cognitive sophistication. Its emotional response suggests that the monkey not only sees the eagle approaching, but also believes that it is an eagle (or takes it in some way to be a particular sort of dangerous animal). Yet we would need more evidence to be convinced that a vervet monkey was actually trying to produce an indication that it believes an eagle is approaching. Absent such evidence, we cannot conclude that the vervet monkey is expressing beliefs by barking.

Imagine that Alan and Bob are spies. Alan dons a bulletproof vest. He knows that Bob is watching, and realizes that his action indicates a belief that he may be shot. Compare three cases. (A) It is okay with Alan that his action indicates the belief that he may be shot, but he is putting on the vest in order to protect himself, not in order to display that belief. (B) Alan is donning the vest in order to display that belief, not in order to protect himself. (C) Alan is putting on the vest for both protection and display purposes. My analysis rules that in (B) and (C), but not in (A), Alan is expressing the belief that he may be shot. This result seems correct. In (B) and (C), but not in (A), it is natural to say that Alan is trying to send a signal to Bob, to communicate with him, to let Bob know what is on his mind. This may seem unclear, but that is to be expected. For (B) and (C) outwardly resemble (A), where there is no expression.

It does not quite suffice to say that S must do something *in order to* provide an indication of belief. Suppose that S wishes to perform an action indicating that he believes $e = mc^2$. To do so, he says "Pass the chalk," walks over to the blackboard, and writes the equation. Then even though S requested chalk in order to display the belief that $e = mc^2$, he did not express that belief until he wrote the equation.[8] The reason seems clear: S intended to display his belief *by* writing the equation, not by requesting the chalk. Imagine next that S routinely depresses the signal lever in order to activate his left-turn indicator. Then even though S intended to provide an indication of the belief that he is about to turn by depressing the lever, he did not mean that he would turn by depressing it. The reason, it seems, is that S did not intend *that act* to be the belief indication. In this routine case, he intended the activation of his tail light to be the indication, not his flipping of the lever. What if the case were unusual in this respect? If S had intended the very act of depressing the

8 Cf. Lombard & Stine 1974, especially p. 211.

lever to be the indication of his belief, then I believe he would have meant by depressing it that he was about to turn.

Let us say that S did A *as* E iff S did A in order to provide E, and moreover intended his doing A to be E. I take this to imply that S intended to provide E by and in doing A. Then we may integrate all of the conditions we have discussed so far as follows:

3.1 **Theorem:** *S expresses the belief that p only if S performs an observable action as an indication that he believes p.*

This principle follows not from previous definitions, but from the final definition of expression to be given later (Definition 3.6). The theorem implies that the expressive action must be observable (as noted in §3.2), but not that its observability must be intended. If people learned to read minds, and could "hear" S's inner speech without his realizing it, S might startle everyone by saying to himself, "You are a bunch of friggers." Others could legitimately ask what he meant by that, I believe, and then accuse him of vulgarity. He could exonerate himself somewhat by insisting, correctly, that he did not intend to express his thoughts. He did not succeed, however, in keeping them to himself.

Together with Definition 2.1, Theorem 3.1 implies that S means that p only if he had a certain intention. In his criticism of Bennett (1976), Harrison objected to the idea that speaker meaning depends on the speaker's intentions.

> If another speaker of English says to me 'Look out, there's a snake next to you,' I precisely do *not* need to undertake any investigation into his beliefs, or his intentions in producing this string of noises, in order to know what he has said. I simply read off the meaning of the sentence he has uttered from the spoken words themselves. (Harrison 1980: 193)[9]

Harrison assumes what I have called the "naive analysis" of speaker meaning in §2.1. It is naive for a number of reasons. First, what the speaker *said* may differ from what he *meant*. If, due to ignorance or a slip of the tongue, S used "entomologist" instead of "etymologist" in describing

9 Cf. Black 1972: §7; Dummett 1973: 149; Pettit 1987: 729; Laurence 1996: 285–6, 292–6. Black goes so far as to claim that intentionalist accounts of speaker meaning are circular because "[i]n standard and unproblematic cases, where we are not dealing with some transposed form of speech such as irony . . ., the relevant speaker's intentions are *constituted* by the meaning of what he says" (1972: 276). Compare and contrast Skulsky 1986: 593 and McKinsey 1978: 191–2, discussed in §4.5. The falsity of the view that Harrison expresses is observed clearly in Grice 1969: 100–1; O. H. Green 1970: 551–2; H. H. Clark 1983; 1993; and Blakemore 1992: 5ff.

Bill, then S meant that Bill is an etymologist even though he said that Bill is an entomologist. And if, as is normal, S happens to utter an ambiguous sentence, then neither what S said nor what he meant can be simply "read off" from the words he uttered. If S says "Sam is an animal," we cannot determine what S said or meant unless we know which Sam, and what sense of "animal," S intended. Even when the speaker uses a sentence that is unambiguous in our native language, we cannot determine what the speaker said or meant unless we know that S is using that language rather than a code in which words of that language have different meanings; and we cannot know that without knowing S's intentions (§7.8). To know what S meant, we will also have to determine whether S was speaking literally or metaphorically. Black's (1972: 276) characterization of metaphorical speech as "nonstandard" does not gainsay the fact that it is a real possibility, one that is actualized with great frequency. It will not help much to refer to "the context" unless the context includes the speaker's intentions. Finally, S did not mean or say anything if he was just testing his voice. And whether he was testing his voice depends on his intentions.

Mark Siebel has argued that Theorem 3.1 conflicts with my analysis of intention. I have argued in Davis (1984a) that S intends to do A only if S expects to do A in part because S wants to do A. When I intend to meet a student, I must expect to meet her in part because I want to meet her (all things considered). Siebel would suggest in the Harrison case, on the other hand, that the speaker expected his utterance to indicate that he believes there is a snake next to you solely because of what the sentence "There's a snake next to you" conventionally means. But any sentence can be used ironically, metaphorically, as part of a code, or with a nonce meaning, and nearly every sentence is ambiguous. Consequently, whether an utterance of "There's a snake next to you" indicates that the speaker believes there is a reptile of the order *Ophidia serpentes* next to Harrison, or a treacherous person next to him, or a plumbing device next to him, and so on, depends in part on what the speaker wants it to indicate. That desire will be part of what caused the speaker to utter the sentence. If the man speaking to Harrison believes that his utterance indicates a belief that there is a reptile next to Harrison, for example, then part of his reason for believing this must be that that is what he wanted it to indicate.

McDowell (1980) and Neale (1992: 545) have objected to accounts of meaning and expression along the lines we are developing on the grounds that the primary purpose of communication is to provide information about the world, not about ourselves. Theorem 3.1 does not

entail, however, that the intention to provide an indication that he believes that p is S's *primary* intention. All that it requires is that S acted with this intention. If I flip the switch in order to turn on the light, my primary intention may be to illuminate the room. But I also intended to flip the switch and turn on the light. Furthermore, McDowell and Neale overlook the fact that speakers do not always intend to provide information about the world when they mean something, as in cases of lying, or of students answering questions.

Vlach (1981: 373) similarly objected that when I give you a reminder, my intention is to make you remember something, not to give you a reason to believe something about me. Vlach's mistake lay in assuming that if S's intention is to do one thing, then S does not intend to do anything else. "S's intention" means not S's *only* intention, but rather S's *main* intention, the intention that is most important among a range of alternatives. Suppose John's intention in going to the nightclub is to pick up a date. We could not infer that he does not intend to drink while he is there. So even when S's (main) intention is to make you remember something, he may also intend to provide an indication of his beliefs.

The same mistake underlies Wettstein's (1991: 74–6) argument that what a speaker refers to does not depend on the speaker's intentions. He imagines an insane historian named Ahern, who believes he is Charles de Gaulle. In a lecture on de Gaulle, Ahern says "And then I marched triumphantly into Paris," meaning that de Gaulle did. Since the point of the lecture was to convey facts about de Gaulle, Wettstein concludes that Ahern did not intend to indicate the belief that *he* marched triumphantly into Paris. But as Wettstein himself observes, Ahern does have this intention as a subsidiary means to his further intention of indicating his belief that de Gaulle marched triumphantly into Paris. So the case refutes only the claim that speaker reference, expression, or meaning depends on the speaker's "primary" intention, something that is not claimed by Theorem 3.1.

Three additional points about the interpretation of "in order to" need to be stressed. First, doing A in order to do B does not require deliberate or reflective action.[10] I can habitually flip the switch to turn on a light, or impulsively smash the ball at my opponent in order to win the point. A child may unself-consciously and without apparent motivation express a belief about something that an adult would be loath to commit himself on without serious deliberation and powerful incentive. Consider the

10 Contrast Evans & McDowell 1976: xx–xxii.

child who blurts out "She's *fat!*" within earshot of a dozen people in the supermarket, and compare the father who is asked whether he agrees. When sincerity is an issue, we tend to regard spontaneous expressions as more reliable signs of belief. Since S has not taken time to weigh the consequences of his act, deception is less likely. When precision is the issue – when we want to know exactly what S believes – we regard deliberate expressions as more reliable.

Second, doing A in order to do B is equivalent to doing A *with the intention* of doing B, which implies much more than doing A *while intending* to do B. It is true, for example, that I intended to marry a woman named Kathy. I knew quite well that she was named Kathy; I also liked the name. I did not, however, marry her *in order* to marry a woman named Kathy. The fact that her name was Kathy was not part of my reason for marrying her. So even if S realizes that buying insurance indicates a belief that driving is risky, and therefore intends to produce an indication of that belief, it would not follow that S bought insurance in order to produce an indication of the belief that driving is risky. Hence it would not follow that S expressed that belief.[11]

Finally, doing A in order to do B does not entail doing B. I can flip the switch in order to turn on the light, even though I fail to turn it on because the bulb is burned out. Hence "S performed an observable action *in order to* provide an indication" does not entail that S *actually* provided one. This is good. For an actual indication of belief is not only insufficient to express it, as the insurance example shows, but also unnecessary.[12] Imagine that S is seated before a panel of buttons, watching objects passing before him one at a time. He is told that pushing the red button indicates that the operator believes the passing object to be red, pushing the yellow button indicates that the operator believes the passing object to be yellow, and so on. S believes what he is told. So when asked later "What color is that one?" he answers by pushing the red button. Then S meant that the object is red. For he pressed the red button in order to produce an indication that he believes the object is red. It does not matter, notice, whether or not pushing the red button *actually* indicates that S believes the object to be red. We can imagine that S was told the truth about the buttons, or a complete fabrication. Either way, he expressed the belief that the object was red.

11 The same principle excludes Grice's, (1969: 152–3) thumbscrew example as well as Black's (1972: 262–3) handshake and gift examples, and distinguishes descriptive from emotive expression (see §3.6, this volume).

12 I missed this point in Davis 1988a, and so used the stronger condition that S *intentionally* performed an action indicating the belief that p. Cf. Recanati 1987a: 121–2.

Because an actual indication is not required, the facts about expression and speaker meaning are sometimes clear when the facts about indication and evidential meaning are not. Consider the following situation:

> On Wednesday morning Ann and Bob read the early edition of the newspaper, and they discuss the fact that it says that *A Day at the Races* is showing that night at the Roxy. When the late edition arrives, Bob reads the movie section, notes that the film has been corrected to *Monkey Business*, and circles it with his red pen. Later, Ann picks up the late edition, notes the correction, and recognizes Bob's circle around it. She also realizes that Bob has no way of knowing that she has seen the late edition. (Clark & Marshall 1981: 12–13)[13]

Now suppose that Ann says to Bob "The movie at the Roxy is silly." Does Ann's utterance indicate that she believes *Monkey Business* is silly (because she knows that *Monkey Business* is playing there), or that she believes *A Day at the Races* is silly (because she knows Bob thinks she thinks *A Day at the Races* is playing there)? This question raises difficult issues, which would have to be resolved by a satisfactory philosophical analysis of "indication," "evidence," "reasons," and related notions.

To determine which belief Ann expressed, however, we do not have to know what, if anything, Ann's actions actually indicated, nor what in general or precise terms an indication is. It suffices to know what Ann *intended* her action to indicate. If she said "The movie at the Roxy is silly" in order to produce an indication that she believes *Monkey Business* is silly, then she expressed that belief. If, on the other hand, she intended to indicate the belief that the other movie is silly, then that is the belief she expressed. If she thought the situation was hopelessly ambiguous, and uttered the sentence with no intention of producing an indication of either belief – just wanting to confuse Bob, perhaps – then she expressed neither belief, and did not mean or imply that either movie is silly. Normally, of course, a speaker would not use a sentence like "The movie at the Roxy is silly" in the Clark and Marshall situation, precisely because the situation is ambiguous.

S's action counts as an *evidential* expression, by contrast, only if it *is* an indication of E. Furthermore, the indication must be reliable. It is self-contradictory to say "The look on Steve's face expresses fear, but Steve is not afraid." Evidential expression, unlike speaker expression, is a species of evidential meaning. A feature expresses fear only if it means fear, that is, only if it means that the subject is experiencing fear. This holds even

13 Quoted in Sperber & Wilson 1986a: 17. See also Blakemore 1992: 19.

for actors. The role of Marian Crane was played by Janet Leigh in the famous *Psycho* shower scene. The look on Crane's face expressed fear *in the movie*. But we could not infer that the look on Leigh's face expressed fear, even though it was an excellent simulation of such a look. For that would imply that she was actually afraid for some reason while the scene was being shot.

Wittgenstein (1953: para. 510) once implied that we could not say "It is cold here" and mean "It is warm here." This is incorrect. We could speak ironically, or use a code in which "cold" means "warm," or simply be mixed up about the meanings of the words. Without something to override the conventional correlation, however, such a feat does seem impossible. It would be hasty to conclude, however, that speaker meaning is not determined completely by speaker intention.[14] For without some overrider, the necessary intention will not be possible either. In a sense, I cannot go to the local grocery store in order to buy a Saturn rocket, because I know I cannot buy one there. In the same way, I cannot use "It is warm" to mean "It is cold" without some stage setting, because I know that without something to override the conventional interpretation, I cannot produce an indication that I believe it is cold by saying "It is warm." Only if a speaker were totally irrational, expecting his desires to be magically satisfied, could he have the indicated intention and mean that it is cold. The same requirement explains why the American soldier could not have meant that he is a German officer by saying "Kennst du das Land, wo die Zitronen blühen?" in Searle's (1969: 44) famous example.

§3.4 SIMULATION

We still lack a sufficient condition of expression. Suppose John and Mary are trying to fool George. Mary pretends that she has been stricken with pain. John rushes over with a great show of concern, deliberately providing an indication that he believes Mary is in pain. While it is correct to say that John *conveyed the idea* that Mary is in pain, he did not *express the belief* that she is. He neither meant nor implied that Mary is in pain. I believe the element of *simulation* is crucial here. John is trying to *hide* his intention

14 Searle (1969: 45), N. L. Wilson (1970: 300–1), and Lewis (1975: 22–3) draw the conclusion. Donnellan (1968: 212), Schiffer (1972: 13), Wright (1975: 372), Carr (1978b: 283), and Martinich (1984a: 122–5) reject it. The root of the problem, I think, is that Searle confused meaning something (speaker's meaning) with meaning what one says (verity); see §2.1. The latter, unlike the former, is partly determined by word meaning.

by pretending to *unintentionally manifest* the belief that Mary is in pain.[15] The typical motivation for such pretense is that an unintentional display is often a more reliable sign of belief than an intentional one. Lying is easy. It is often harder to fake an action that in the normal course of events unintentionally manifests a belief. Paradigm cases of communication contain no pretense or deceit at all.

Here is a test case. Suppose that Alan asks Bob whether driving is risky. Bob says "Watch" and proceeds to buy auto insurance. Then Bob bought insurance as an indication that he believes driving to be risky. And even though buying insurance is normally an unintentional sign of that belief, Bob was not *pretending* to provide such a sign. Did Bob express the belief that driving is risky? I believe he did, doing so effectively if unconventionally. Indeed, we might imagine Alan replying "I disagree" and then tearing up his own policy. It is natural to say that Bob successfully attempted to send a signal to Alan, to communicate with him, to let Alan know what was on his mind, specifically, to indicate (in the speaker sense) that driving is risky. By buying insurance, he meant or implied that driving is risky.

"Pretense" has a strong and a weak sense. When Jacqueline Bisset played the role of Jackie Kennedy in a movie, she did her best to look and act the part. We would not say, however, that she was pretending to be Jackie Kennedy. For there was no *intent to deceive*, hence no *dissembling*. Bisset was not trying to pass herself off as Kennedy. By contrast, when a child says "Let's pretend" and then goes on to imitate a cat, we would say that he was pretending, even though there was no attempt to fool anyone. And the director of a later movie might try to elicit a particular effect by telling Bisset, "Just pretend you're playing Jackie Kennedy." We shall use the term "simulation" generically, with no implication of dissembling. A simulation designed to be incorrectly taken as genuine is *covert*, let us say, otherwise it is *overt*. "Pretense," "feign," and "fake" imply covert simulation in their strong sense.

Meaning and expression exclude covert but not overt simulation. To see this, compare the following cases. (A) Stu is trying a new dish prepared by a friend. Upon tasting the food, Stu immediately contorts his face and spits it out in an attempt to perfectly fake a reflexive response, making his

15 Cf. Strawson 1964: 446, 454, 460; Bach & Harnish 1979: 153; McDowell 1980: 128; Blackburn 1984: 114–18; Martinich 1984a: 115; Sperber & Wilson 1986a: 30–1, 60–1; Kemmerling 1986: 147, Recanati 1986: 226–34, Bertolet 1987: 205, Avramides 1989: 50; Christensen 1997: 502–6. The example came from a reader many years ago.

friend think he found the dish to be awful. (B) When asked "How is the food?" Stu contorts his face and spits out the food in an obvious imitation of a reflexive response. In both cases, let us stipulate, Stu is simulating a reflexive response in order to produce an indication that he thinks the food is awful. It seems evident that in (B), where the simulation is overt, Stu does mean something by his action, and is nonverbally expressing a belief. It seems equally evident that in (A), where the simulation is covert, Stu does not mean anything by his action, and is not expressing a belief.

Let the relation between S and the belief that p described by the consequent of Theorem 3.1 be called *conveying*. Then Theorem 3.1 may be strengthened as follows:

3.2 **Theorem:** *S expresses the belief that p only if S conveys the belief that p without thereby covertly simulating an unintentional indication of the belief.*

For short, let us say that S must *undisguisedly* convey that belief. Alternatively, we will say that S provided an *undisguised indication* that he believes p. Theorem 3.2 should be understood as allowing S to express a belief *while* pretending to unintentionally convey it, as long as he also conveys the belief undisguisedly by doing something else. Thus John could express the belief that Mary is in pain by saying "Mary is in pain" while rushing over to her with a phony show of concern. His utterance satisfies the definiens of Theorem 3.2 even though his show of concern does not. Let us also stipulate that a simulation of an unintentional belief indication counts as overt as long as the speaker intends someone in his audience to recognize that it is not genuine. If Stu intends his wife to recognize his act, he can be said to have meant that the food is awful, even if he does not intend his host to recognize it.[16]

§3.5 OCCURRENCE

The conditions formulated so far almost suffice to define expression for belief. As noted in §2.2, we can believe that $2+2=4$ even when we are not currently thinking that thought. Nevertheless, we cannot express the belief that $2+2=4$ without expressing the thought that $2+2=4$. Doing the former is a way of doing the latter. Our definition should therefore yield this as a theorem.

16 I am indebted here to Christian Plunze.

3.3 **Theorem:** *S expressed the belief that p only if S expressed the thought that p.*

Uttering grammatical *sentences* is the normal means of expressing our thoughts. The most common way of expressing our beliefs is to utter *declarative sentences in unmarked contexts*, those in which the sentence is not part of a compound sentence or a work of fiction, not accompanied by a gesture of irony or jest, and so on. Thus uttering "Mary will sing" indicates in any context that S is thinking the thought that Mary will sing, assuming that S knows English. Its utterance in unmarked contexts also indicates that S believes that Mary will sing. Cases in which "Mary will sing" is uttered in jest or irony show that the converse of Theorem 3.3 fails.

Theorem 3.3 can be secured by requiring that to express a belief, S must display *occurrent* belief. S occurrently believes p provided that S both believes and is thinking the thought that p (§12.5). When Theorem 3.2 is strengthened accordingly, the result is our definition of expression for belief.

3.4 **Theorem:** *S expressed the belief that p iff S performed an observable action as an undisguised indication that he occurrently believes that p.*

Different forms of "express" are definable in the same way by varying the form of "perform." For example, S *expresses* a belief, either frequently or on a particular occasion, iff S *performs* an observable action as an undisguised indication that the belief is occurrent.

Definition 2.1 and Theorem 3.4 imply, of course, that S means that p only if S undisguisedly conveyed the occurrent belief that p. S means that p specifically *by (producing) e* only when S expressed the belief *in virtue of producing e*, that is, when the observable action S intended to be the belief indication was producing e. Since we are focusing on the exclusive sense of speaker meaning, which requires direct expression, we have:

3.5 **Theorem:** *S meant that p by (producing) e iff S produced e as a direct and undisguised indication that he occurrently believes p.*

We may similarly define S *used e to* mean or express the belief that p. Since by Definition 3.6 a belief is expressed if and only if it is expressed in virtue of performing some action, Theorem 3.5 and Definition 2.1 entail that S meant that p iff S meant that p by some e. As for cogitative speaker meaning, we get: *S means μ by (expression) e iff by producing e S directly and undisguisedly conveyed the occurrent idea μ.*

Theorem 3.3 suggests that cognitive speaker meaning implies cogitative speaker meaning – that *S means that p only if S means "p."* While appealing, this implication fails. In the exclusive senses we are concerned with, cognitive and cogitative meaning involve the *direct* expression of beliefs and thoughts, respectively. And the direct expression of beliefs allows more indirection than the direct expression of thoughts (see §2.4). Steve may mean that he is in great financial shape when he uses "I am in hock" ironically. But Steve does not mean "I am in great financial shape" by *the sentence* "I am in hock." For Steve expressed the thought that he is in great financial shape by expressing the thought that he is in debt. Steve's utterance was intended to indicate the thought that he is in debt *and thereby* the thought that he is in great financial shape.

§3.6 GENERAL DEFINITION

Expression can be defined for mental states other than belief in the same way. Thus "belief" can be replaced by "desire," "intention," or "hope" in Theorem 3.4. The same pattern of definition even holds for the expression of occurrent states, such as thought. To express the thought that $2 + 2 = 4$, S must provide an undisguisedly intentional indication of the occurrent thought that $2 + 2 = 4$. When applied to thinking, however, "occurrently" is redundant. Thinking is necessarily occurrent. So to convey the thought that p is to convey the occurrent thought that p. Similar remarks apply to ideation in general. To express the idea of a dog is to provide an undisguisedly intentional indication that the idea is occurrent.

The most obvious way to generalize Theorem 3.4 is to say that S expresses M iff S performs an observable act as an undisguised indication of occurrent M, where "M" stands for the name of any mental state.[17] There are cases, however, in which the right-hand side of this definition holds but not the left. Let "M" be "Frank's favorite belief," for example. Now suppose that A asks "Is Frank's favorite belief occurring to you?" and S answers "Yes," even though he has no idea what Frank's favorite belief is. Then S's answer was an observable act, and he performed it as an undisguised indication that Frank's favorite belief was occurring to him. It does not follow that S expressed Frank's favorite belief, unless that (strangely) happens to be the belief that Frank's favorite belief is occurring to S. I suggest that doing something with the intention of indicating the

17 Cf. Davis 1988a: 282.

occurrence of Frank's favorite belief is not the sort of intention involved in expressing a belief, because it contains the descriptive concept "Frank's favorite belief." The reason for this, I believe, is that this sort of concept is not *introspectively applicable*. That is, we cannot tell on the basis of introspection alone that Frank's favorite belief is occurring to us. We can know this only if we know on the basis of information about Frank what his favorite belief is and know introspectively that that belief is occurring to us. It is not enough to note that we can only express introspectible states. For Frank's favorite belief is the belief that God exists, and that is introspectible. We must also note that expression requires an intention containing a "mode of presentation" of the state that can be applied in the process of introspection.

So let "Ψ" stand for terms expressing introspectively applicable concepts. Since only mental states are objects of introspection, it follows that "Ψ" stands for terms denoting mental states. Then Theorem 3.6 can be generalized as follows.

3.6 **Definition:** *S expresses Ψ iff S performs an observable act as an indication of occurrent Ψ without thereby covertly simulating an unintentional indication of Ψ.*[18]

In addition to occurrent states and propositional attitudes, we are also said to express *propositions.* While on many views propositions are not mental states, I define them as thoughts (see Chapter 13). So Definition 3.6 applies directly. If propositions are defined differently, expression can be defined for them in one of two ways. We could use "S expresses the proposition that p if and only if S expresses the thought that p" as a definition, taking the definiendum to be an idiom. Or we could use Definition 3.6 after allowing "Ψ" to stand for names of propositions and stipulating that an occurrent proposition is one that S is thinking.

Terms that do not meet the restriction on "Ψ" include "the knowledge that p" and "an unconscious desire for Φ." We can provide indications that such states are occurrent, but we cannot be said to express them. We can express belief, including the belief that we have knowledge, but not knowledge. There are some introspectible mental states that we do not

18 As explained in §6.2, we are taking "Ψ" to occupy an *opaque* context on both sides of Definition 3.6. But either side can be given a transparent interpretation. Thus if S has expressed the belief that God exists, then there is a sense in which S has expressed Frank's favorite belief. Definition 3.6 can appear to fail if we equivocate and give one side an opaque interpretation and the other a transparent interpretation.

speak of expressing,[19] though, for which there appears to be no good reason to deny that we can express them. It is easy to imagine a man doing something as an overt indication that he has a headache. He might, for example, utter the sound "anhk," for example, or draw a picture of a head with a little hammer hitting it, or say "I have a headache." We do not speak of expressing headaches. But there seems to be no reason to deny that we are doing so in the cases imagined. The cases seem to be completely analogous to those in which we express pain or irritation.

Definition 3.6 holds straightforwardly for emotions such as joy and fear. Koch (1983) has distinguished *expressing* an emotion (e.g., saying "Damn you!" or "I'm mad at you" to express anger) from two other ways of manifesting one: *evincing* (turning red in anger) and *coping* (trying to kill the person one is angry at, or trying to calm down). Evincing differs from expressing in being an involuntary effect of the emotion.[20] Coping is intentional behavior, but it is aimed at dealing with the object of the emotion in some way, or at controlling the emotion itself. Coping behavior is not typically intended as an indication of the emotion. The spy and insurance examples show, however, that coping and expressing are not incompatible. Alan might buy insurance both in order to avoid vulnerability to the harmful financial effects of having an accident, and in order to display a fear of having an accident. Definition 3.6 seems adequate, therefore, to distinguish expressions of emotion from other manifestations.

Green (1970: 56) and Hampshire (1972: 154) claim that expressing a belief differs from expressing an emotion in requiring the use of language. It is true that emotions can be expressed nonverbally. Thus we can express anger by clenching our fist in a symbolic gesture, or joy by throwing our hat in the air. But belief can also be expressed nonverbally. If someone asks "Who won?" S can express the belief that John did by pointing to John, or even by slapping him on the back. If someone is trying to start his car in vain, S can express the belief that the car is out of gas by holding an empty gas can upside down. In both cases, S is intentionally performing an action that indicates that he has a belief, and is not doing so by pretending to unintentionally manifest the belief. The verbal-nonverbal distinction may have seemed overly important because it was mistaken for that between speaker and evidential or natural expression.

19 Cf. Tormey 1971: 5.
20 Cf. Benson 1967: 339–40.

There is a conspicuous distinction between expressing anger by saying "You bastard!" and doing so by saying "I am angry."[21] The same distinction exists between expressing pain by saying "Ouch!" and doing so by saying "I am in pain." Let us call the former mode of expression *emotive*, the latter *descriptive*. One difference is that emotive expression is typically spontaneous, whereas descriptive expression is typically deliberate. But these attributes are sometimes reversed. Emotive expression is exclamatory, but descriptive expression may be too. The essential difference lies in the fact that a subject who says "You bastard!" expresses anger but not the belief that he is angry. By contrast, a subject who says "I am angry" expresses anger *by* expressing the belief that he is angry. Hence someone can mean, say, and assert that he is angry by saying "I am angry" but not by saying "You bastard!" Of course, given the power of introspection, saying "Ouch!" indicates not only that S is in pain, but also that S believes that he is in pain. Nevertheless, a speaker usually says "Ouch!" in order to display pain, not in order to display the belief that he is in pain. So by saying "Ouch!" S expresses pain but not a belief about pain.

It follows that if S does not believe he is angry, then S is lying in the descriptive case but not in the emotive. Since he is insincere in either case, we may conclude that lying is a special form of insincerity requiring the expression of belief. Similarly, if S is not angry, then he utters a falsehood when he says "I am angry" but not when he says "You bastard!" The subject can express anger emotively by expressing a belief, as when S expresses anger by screaming "That's the worst thing you've ever done!" But S is not here expressing the very belief that he is angry, and so does not mean that he is. S might be lying, but not because he knows he isn't angry.

There are other nonemotive ways of expressing emotion. One can express anger by saying "You are going to pay," although without some contextual or verbal clue this would more naturally be interpreted as a simple statement, expressing belief without expressing emotion. One has expressed anger indirectly but not descriptively – by expressing a belief other than the belief that one is angry. One can also express anger by using an interrogative sentence ("Didn't you see the stop sign?") or an imperative ("Come here right now!"), usually in a special tone of voice.

21 This is discussed in Meinong 1910: 25–7; Alston 1965; Benson 1967: 336, 338, 347–9; O. H. Green 1970: 563–5; Tormey 1971: 7; Black 1972: 262–3; and Wierzbicka 1992b: 162–3. See also §2.5. Alston missed the fundamental difference noted in the following, and therefore minimized the distinction.

Benson (1967: 338) observed that an emotion can often be expressed emotively when the linguistic resources to do so descriptively are lacking. This is one value of poetry.

Wierzbicka (1992b: 162–3) suggests that the difference between descriptive and nondescriptive expression is that only the former has illocutionary force or purpose. Whereas "I am in pain" implies "I say that I am in pain," "Ouch!" does not. It is true, of course, that the utterance of "I am in pain" will make "I say that I am in pain" true in a typical speech context, where the speaker is not just testing his voice or giving a linguistic example. It is also true that a speaker who utters "I am in pain" in the typical speech context will normally believe "I say that I am in pain." But "I say that I am in pain" is not part of the meaning of "I am in pain." Someone can perfectly well think or believe "I am in pain" without thinking or believing "I say that I am in pain"; the truth of "I am in pain" does not entail that anyone says anything. Furthermore, a man who expresses pain by saying "I am in pain" need not be expressing the belief that he is saying anything. He need not in any sense be talking about the speech acts he is performing. So while the utterance of a sentence may imply that the speaker says something, the speaker need not be implying that he is saying something.

4

Alternative Analyses

Grice originated the program of defining speaker meaning in terms of intention, and provided what is still the benchmark analysis.[1] Many variations have been proposed, all fairly close to the original. I shall review a wide variety of facts that show that Grice's basic approach was wrong, and that mine is more promising. While my analysis is a descendent of Grice's, the fundamental conception has been transformed. Speaker meaning is not the intention to communicate. It contrasts with other semantic acts in that it need not be audience-oriented. Speakers can mean something without expecting or hoping that their intentions be recognized.

Most of this chapter will focus on cognitive speaker meaning, comparing Grice's analysis to Theorem 3.5, according to which S meant that p by e provided that S used e as an indication that he occurrently believes p, without covertly simulating an unintentional indication of the belief. In light of the equivalence between meaning that p and expressing the belief that p (Definition 2.1), the discussion will be an implicit comparison of a Gricean analysis of expression with the analysis provided in Chapter 3.

1 See Grice 1957, 1968, 1969, 1982, 1986. Grice's basic approach, has been picked up in Strawson 1964, 1971; Stampe 1968; Patton & Stampe 1969; Searle 1969, 1979, 1983, 1986; Armstrong 1971; Bennett 1971, 1973, 1976; Schiffer 1972, 1982, 1987a, 1992; Walker 1975: 155; Kempson 1975: 138; Holdcroft 1978; Platts 1979; Schwarz 1979: xxxi-iv; Loar 1976a, 1976b, 1981; Martinich 1984a; Blackburn 1984; Suppes 1986; Avramides 1989; 1997: §§6–7; Chierchia & McConnell-Ginet 1990: 148–57; Neale 1992: §5; and Meggle 1997. The Gricean approach has been taken in somewhat different directions in Lewis 1969; Vendler 1972: 62–3; Schwarz 1979: Chapters 2–4; Bach & Harnish 1979; Hungerland & Vick 1981; Kemmerling 1986; Sperber & Wilson 1986a, 1987; Recanati 1986, 1987a; Bach 1987a, 1987b, 1994a, 1994b; Devitt & Sterelny 1987: §7.4; Bertolet 1987; Thomason 1990: 342–4; Schiffrin 1994; and Christensen 1997: 506. A similar view was developed by Anton Marty (1908), a student of Brentano. See Liedtke 1990.

The major extant competitor to the Gricean analysis will be critiqued in §4.4. This approach seeks to define speaker meaning or expression in terms of commitment and truth rather than in terms of intention and indication. Commitment has been defined variously, in terms of lending one's authority to a belief, or of conventional norms requiring belief, or of intended truth or verification conditions.

The few extant analyses of cogitative speaker meaning will be criticized in §4.5. Whereas Grice focused on "parts" of cognitive speaker meaning, others highlighted intentions concerning word meaning. The defeatism that has clouded the literature on speaker meaning can be lifted, I argue, by making different choices in the execution of the Gricean program.

§4.1 PRODUCTION OF BELIEF

According to the main condition of the original Gricean analysis, what a speaker means is determined by what belief he intends to produce in his audience:

(1) S means that p only if S intends to produce in some audience A the belief that p.

This condition holds in a wide variety of cases. For example, when I said earlier "We shall focus here on cognitive speaker meaning," I intended to produce in you the belief that we would do just that. One of the standard purposes of expressing a belief is to produce that belief in one's audience. Nevertheless, condition (1) also fails in a wide variety of cases.[2] For example, adults rarely if ever try to produce the belief that $2 + 2 = 4$ in one another, since everyone knows that everyone already believes it. Nevertheless, we often say and mean that $2 + 2 = 4$. A reminder typically presents a similar problem, as does a student's answering a teacher's question and a wife's telling her husband that he is lying.[3] S need not intend to produce a belief even when A is known to lack it. The sheriff,

2 For a review, see Black 1972 and Vlach 1981. See also Schiffer 1972; 1987a: Chapter 9; and Avramides 1989, which defend Grice's general approach.

3 These last two cases show that the replacement of "belief" with "occurrent belief" (Schiffer 1972: 45, 51; 1987a: 247) or "entertained belief (thought, proposition)" (Neale 1992: 547) is still too strong. Incredibly, Schiffer says, "I am inclined to think that none of these cases present a serious problem and that the sense in which it may be said of S in these examples that he meant that p . . . is an extended and attenuated sense . . ." (1972: 71). No such ambiguity is detectible, at least to my ear.

who dutifully told Mrs. Hubbard that her son is dead, may have known, and been glad, that she would refuse to believe him.[4] If S knows that he is mistrusted, or will be misinterpreted, S may actually intend to make A believe the opposite of what he means.[5]

Grice (1968: 230; 1969: 171–2) later suggested meeting such counterexamples by requiring only that S intend to produce in A *the belief that he (S) has the belief that p*. Standardly, a speaker intends his audience to engage in the following process of reasoning:

S said "p."
∴ S believes that p.
∴ p.

The audience is thus standardly expected to acquire the belief that the speaker believes p on the basis of what the speaker said, and to acquire the belief that p on the basis of the fact that the speaker believes p. But as the cases in the previous paragraph show, the speaker often does not expect, or even want, the second inference to be drawn, while still intending the first to be drawn. So Grice's revision has the effect of weakening his analysis. In particular, the revision remedies the student–teacher case, and cases of countersuggestion with a speaker believed to be truthful but mistaken about the facts. To handle common knowledge, "occurrent" must be added before the first or second "belief." When I say and mean that Russia is a large country, I may intend to make occurrent your belief that it is, or your belief that I believe that it is; but I surely will not intend to give you either belief, unless I take you to be a child or an ignoramus.

Still problematic are those cases of countersuggestion, expected refusal to believe, and reminders in which S is thought to be a liar. So are answers to rhetorical questions. When I ask a rhetorical question like "Is the Pope Catholic?" I expect you to already know the answer and to know that I know it, since that is what makes the question rhetorical. And since I just asked the question, I expect your knowledge to be occurrent. So when you go ahead and answer "Yes, the Pope is Catholic," you cannot expect

4 See also Carr's (1978b: 283–4) example of a suspect objecting that he was elsewhere at the time of the crime, Harman's (1974: 128–9) of asserting what you believe in the presence of those forcing you to recant, Chomsky's (1975: 61) of protesting something to the IRS, knowing "perfectly well that the 'reader' (probably a computer) couldn't care less," and Alston's (1982: 262) of prefixing one's remarks with "I know you won't believe this, but. . . ."

5 See Carr's (1978b: 285) example of orders given with the intention that they be disobeyed, Kemmerling's (1986: 139) oracle case, and especially Black's (1972: 265–6) "truthful liar."

to inform me either that you believe that the Pope is Catholic, or that you occurrently believe it.

The rhetorical answer problem can be eliminated by replacing the intention to *give A the belief* that S believes p with the weaker intention to *give A a reason for the belief*, which was Armstrong's (1971) suggestion.[6] This is very close to the intention I require for meaning and expressing the belief that p. Indeed, a speaker normally produces an indication that he believes p in order to give his audience a reason to believe that he does. As a natural sign of the belief, the action would be the reason. When the defendant maintains his innocence in court despite knowing that everyone believes he is lying, he does not expect to make anyone believe that he is innocent. But he does intend to provide an indication, and give the jury one reason to believe, that he is innocent.

Armstrong's condition is still too strong, however. First, in the Clark and Marshall situation (§3.3), Ann may have said "The movie playing at the Roxy is silly" as an indication that she believes *Monkey Business* to be silly, knowing full well that her utterance would not indicate that belief to Bob. Indeed, Ann may think that she will be giving Bob a reason to believe that she thinks *A Day at the Races* is silly. Second, travelers in foreign lands often speak their own language hoping, but not intending, that they will be understood. More radically, a speaker may deliberately say and mean that p *in a language known to be unknown to his audience*, as when a professional athlete curses in his native tongue so that the spectators won't be offended.[7] Such an utterance indicates that the speaker believes p, though not, he realizes, *to his audience*. Since he knows that they do not know how to "read" the sign, he does not intend it to give them a reason to believe what it indicates. Finally, S's audience may be *unperceptive*, *unintelligent*, *unconscious*, or even *dead*. It is not altogether rare for people to speak to babies, pets, or the recently departed. But few intend to give them reasons to believe anything.

Such substandard audiences show more generally that S need not intend to have *any* impact on those to whom his remarks are addressed.[8] Indeed, the bare requirement that S *have* an intended audience is inappropriate

6 Cf. Bach & Harnish 1979: 15; Holdcroft 1978: Chapter 8; Bach 1987a: 51; 1994b: 18, fn. 17; and Recanati 1986: 218–22. Sperber and Wilson (1986a: Chapter 1) require in a similar vein that S intend to *make the belief manifest to A*. Hobbes (1655: Chapter 2) said that a name is a sound uttered to others so that it might be a (natural) sign *to them* that a certain thought either occurred or did not occur to the speaker.

7 Cf. Ziff 1967: 446; Harman 1974: 229; and Elugardo 1997: 178.

8 Cf. N. L. Wilson 1970: 296–7. Contrast Wright 1975: 364, fn. 6.

in an analysis of speaker meaning, implication, or expression.[9] Schiffer (1972: 73–7) described several cases, among them recording something in a diary, and writing furiously in private just to solve a problem. I would add venting one's frustrations by screaming "You are a ____" precisely *because* no one can hear. While Schiffer thought otherwise, these seem as clear as any cases of speaker meaning. But with no intended audience, S cannot be intending to give one any beliefs or reasons nor to elicit any response therefrom. Of course, S is sometimes "talking to himself."[10] But not in every case. And unless S is mentally ill, the main Gricean condition will hold when A is S himself only if S intends to review later what he is saying now. But this intention is not essential, as the frustration case illustrates.

The fundamental defect of the Gricean analysis is that it is *audience-oriented*. Many speech acts do require an audience, such as *communicating*, *telling*, and *informing*. We cannot tell, for example, unless we tell someone, and we cannot communicate unless we communicate with someone. But meaning, implying, and expressing are different, as are saying, signaling, and indicating. "S meant something" does not require, or even allow, completion by "to ____." Applications of the Gricean analysis to communication, telling, and informing are thus more promising than Grice's own application to speaker meaning.[11]

Grice later suggested replacing the reference to a specific audience A with a general reference to *suitable* audiences.[12]

(2) S means that p only if S intends to produce in any suitable audience the belief that he (S) believes that p.

9 This has been noticed in O. H. Green 1968; 1970: 558; Tormey 1971: 92–3; and Koch 1983: 180 in the case of expressing, and in Black 1972: 264; Chomsky 1975: 60–4; Harman 1977a: 422; Yu 1979: 278; and Lawrence 1996: 277–8, 297 in the case of meaning. Describing cases of meaning without an audience as "noncentral," "secondary," or "derivative" does not eliminate the problem, as Bennett (1976), Suppes (1986), and Bertolet (1987) seem to maintain. Cf. Hungerland & Vick 1981: 43.

10 Cf. Searle 1986: 211 and Avramides 1989: 66. When discussing audienceless cases, Griceans seem to shift, without awareness, from the standard conception of an audience as the individual or group to whom one is talking and with whom one wishes to communicate, to the generally broader notion of the individual or group that can hear and understand what one is saying.

11 See Stampe 1968: 141; Black 1972: 277–8; Chomsky 1975: 68; Wright 1975; Hacking 1975: 22; Bach & Harnish 1979: Chapter 1; Yu 1979; Searle 1983, 1986; Kemmerling 1986; Sperber & Wilson 1986a: Chapter 1; Recanati 1986; and Chapter 5 of this volume. See also Landau 1984, 215.

12 See Grice 1969: 174–7; Schiffer 1972: 75–6; Hyslop 1977; Hungerland & Vick 1981: 40–1; and Bertolet 1987. Contrast Chomsky 1975: 67–8.

S is allowed to have in mind only a general description of a potential audience, such as "passerby" or "speaker of English," and an actual audience automatically qualifies as suitable, Grice says. But S may realize that his actual audience is a cat, and that some speakers of English who pass by would not understand him or take him to be serious. Alternatively, S may not *care* what his actual or potential audience would believe. Equally seriously, there will often be no one common effect that one wishes to produce in all suitable listeners, no matter how "suitable" is defined. The stereotypical politician, for example, might utter a vague generality like "We must support quality education for our children," intending conservatives to form the belief that he supports government funding for private schools, and liberals to form the belief that he favors increased federal funding for public schools.[13] But unless S intends to produce the same belief in *every* suitable audience, (2) will rule incorrectly that S means nothing.

Hungerland and Vick (1981: 40–1) have claimed that an individual cannot *signal* unless he takes there to be a real possibility of an audience. Their example of a shipwrecked sailor sending out an SOS makes this claim quite plausible. If, unbeknownst to the sailor, he is the last person left in the entire universe, we can still count him as signaling. But if we imagine that the sailor somehow *knows* that he is alone in the world, then we cannot. He might tap out "SOS," of course. But we could not describe him as having signaled anything. Signaling appears to imply at least the *hope* of communicating. This point can be granted, since signaling is more specific than meaning. The relevant conclusion to be drawn is that belief in the real possibility of an audience is one of the features distinguishing signaling from other species of meaning. For even in the case we have imagined, the sailor could perfectly well *mean* that he has no food by the dots and dashes spelling out "No food." We can imagine him "talking" to himself using the telegraph key, or writing "I am alone in the world" in his diary. He may have meant what he wrote or keyed, even though we could not say that he signaled anything.

On my view, cognitive speaker meaning is essentially an attempt to produce an indication of belief, in a certain way. Since indicators do not require an audience, the real or perceived absence of one does not prevent meaningful speech. While we might occasionally wonder why, speakers without audiences sometimes do make sounds intended to indicate

13 See also the Meese example in Bertolet 1987: 208 and the wedding example in Martinich 1984a: 122.

beliefs. Many have wondered why, on my account, people would bother to provide indications of their beliefs when their only audience is a baby. And they have suggested that the lack of a reason would be evidence that people have no such intention, undermining my account.[14] First, I reject the rationalistic thesis that people intend to do something only if they have a reason for intending to do it. I believe many actions, from doodling to enjoying oneself, provide counterexamples. We often intend to do things because we "just feel like it," and talking to babies is a good example. Second, it is agreed by all that we intend to talk to babies. If we have to have reasons for things we intend to do, then whatever reasons we have for talking to babies could be our reasons for intending to provide indications of our beliefs in their presence. One might suggest, for example, that we talk to babies so that they can learn to speak. I do not think this is plausible, but my point now is that it is no less credible to suggest that we utter words in the presence of babies with the intention of indicating our beliefs for the same reason, since learning our language requires learning what thoughts and beliefs we use words to express.

Husserl (1900: 279) maintained that an indication is necessarily an indication *to* someone. Even if this were true, it would not follow that if S intended to provide an indication of something, then S intended to provide an indication of something to anyone in particular. One cannot get married, for example, without getting married to someone. But many people intend to get married even though there is currently no one in particular they intend to marry. Furthermore, we noted in §3.2 that Husserl's thesis is false. A patient's symptoms may indicate that he has Lyme disease even though no one realizes that they do. Indeed, the indications of Lyme disease are often misdiagnosed.

Given the well-known difficulties with definitions of related concepts like causation, evidence, and probability, it would be remarkable if indication were semantically definable. Rough glosses are possible, of course. "A indicates B" says roughly that there is a causal or statistical relation between A and B in virtue of which A would give a suitably placed observer a reason to expect B. It would follow from such a definition that intending to produce an indication of B implies intending to give a suitably placed observer a reason to expect B only if the definiens and definiendum are exact synonyms, which is highly unlikely. Condition (2) might hold, though, for "ideal" speakers, who have fully mastered the

14 This objection was pressed by Mark Siebel, Christian Plunze, Christoph Jäger and others at the University of Leipzig.

concepts involved in their intentions. It might also hold true if a suitable audience is defined as something like *an audience ignorant of S's beliefs that would correctly interpret indications thereof.* But I see no reason to believe that a speaker would *have* to have an intention concerning such audiences if he were to mean anything.

§4.2 RECOGNITION OF INTENTION

Grice's (1957) analysis had an auxiliary condition: S must intend to produce in A the belief that p *by means of A's recognition of his (S's) intention*. Suppose Mason secretly drops Susan's handkerchief so that the detective will believe that Susan was the murderer. While Mason intended to produce that belief in the detective, Mason did not mean by his action that Susan was the murderer. The reason, according to Grice, is that Mason did not intend the detective to recognize his (Mason's) intention to produce that belief. When Henry Cabot Lodge showed a photograph of Soviet missiles in Cuba to convince a skeptical UN, he intended everyone to recognize that he intended to produce belief. But Lodge did not thereby mean that there are Soviet missiles in Cuba, on Grice's view, because he did not intend his intention to do the convincing.

On my analysis, Mason meant nothing because he did not intend his action to indicate anything about his beliefs. The reason *cannot* be that Mason did not intend anyone to recognize his intention. For such an intention is unnecessary for speaker meaning. This is shown by all of the previously discussed cases in which S means that p but has no intention of producing the belief that p (or any related belief) in an audience, such as when S is deliberately speaking a language that his audience does not understand.

Meaning without intended recognition of intention is possible even when the speaker does intend to produce belief. Suppose that in order to get Alan to realize that his wife is having an affair, I write him a note to that effect *without signing it* (perhaps even addressing it to someone else), knowing that Alan will believe the note as long as he does not think that it is from me. I do not intend him to recognize *my* intention. I may not even care whether or not Alan recognizes that *the author (whoever he is)* intends him to conclude that his wife is having an affair. For I may think that even if Alan dismisses the author as a crank, mere consideration of the possibility of infidelity will open Alan's eyes to the evidence plainly before him. Nevertheless, I meant what I wrote. The conditions of Theorem 3.5 are satisfied: I intend the note to indicate my belief, but not to Alan.

S might also realize that recognition of intention will play *no role whatsoever* in the production of belief. If I say "I'm over here" in a crowd to let you know where I am, I mean what I say even though I expect you to locate me on the basis of sound alone. And Herod might have said and meant that John the Baptist was dead when presenting his head to Salome, knowing that she would reach that conclusion without even hearing what he said.[15] Ambassador Lodge may even have meant something by showing the picture. Just imagine that someone asked "What kind of missile did the Soviets put in?" And imagine that Lodge, not knowing the name, answered by holding up the picture. The agents intend in such cases to provide an indication of their belief, so my analysis can be satisfied.

As I see it, recognition of S's intention is necessary only to *understand S*, that is, to understand what S means. While you can understand the speaker's *words* without such knowledge, you cannot understand *the speaker* without knowing what his intentions were in uttering those words. To see this, suppose that the speaker says something like "He flew to the bank," or imagine a scenario in which the speaker might be just reading a script or testing his voice.[16] Since S cannot *communicate with A* unless A understands S, A must recognize S's intentions if S is to communicate with A. So, for a second reason, a Gricean analysis is more promising for communication than for speaker meaning. Intending an audience to recognize your intentions and understand you is not necessary for meaning something. Indeed, S might deliberately baffle his audience by using an unfamiliar language, a private code, or some impenetrable jargon. The widespread assumption that meaning something involves attempting to communicate, which seems to have propelled the Gricean analysis, is false.

§4.3 HIGHER-ORDER INTENTIONS

Grice later concluded that his original analysis was too weak on the basis of examples presented by Strawson, Stampe, and Schiffer involving what I call *semi-covert simulation*.[17] In the Stampe example (Grice 1969: 154), Jack is playing cards with Jean (his boss), and wants her to know that he

15 Cf. MacKay 1972: 65–6; Vlach 1981: 379; Recanati 1986: 275; and Neale 1992: 548–9. Contrast Bach 1987b.

16 Contrast N. L. Wilson 1970: 299–302; Black 1972: 276; and Blackburn 1984: 113. Cf. Dummett 1973: 149; Sperber & Wilson 1986a: 10–1; and Bertolet 1987: 200.

17 See Strawson 1964: 446–7; Grice 1969: 154–9, 1982: 237–43; Schiffer 1972: 17–22; 1987a: 245; Bennett 1976: 126–7; Avramides 1989: 48–50; and Christensen 1997: 506.

has a good hand without letting her know too blatantly. So Jack simulates pretty closely a smile of pleasure, expecting Jean to recognize that it is fake, while nevertheless expecting her to think that he intends her to take it as a spontaneous giveaway. Grice thinks that Jack did not mean by smiling that he had a good hand, even though his original conditions are satisfied.

Grice therefore suggested adding either a second-order intention clause, or a prohibition clause. The literature has generally taken the former option, requiring that S *intend A to recognize his first-order intention*, viz., the intention to produce the relevant belief in A by means of recognition of intention. Some have proposed requiring that S intend A to recognize his nth-order intention *for every n*.[18] The second-order intention clause excludes the Stampe example, since Jean is intended to think that Jack intends to produce belief by means *other* than recognition of intention.[19] But it also excludes all of the genuine cases of meaning excluded by the first-order intention clause, such as those in which people recite common knowledge, deliberately speak a language unknown to their audience, write unsigned notes, soliloquize, say "I am here," or use photos to express their beliefs.

The higher-order intention clauses added to the Gricean analysis have also led to the "psychological reality" objection: the intentions required seem to be too complex for all speakers at all times – indeed, for any.[20] As Loar put it, how can people intend things that they are incapable of even thinking? How could speaker meaning ever be recognized if an infinity of intentions is required (or excluded)? Such an objection has little force against my analysis. It is not implausible that all speakers intend their utterances to be indications of certain thoughts or beliefs, without covert simulation (or semi-covert), when they mean something. If it is doubtful that an individual has such intentions, as with a ten-month-old girl who makes a noise sounding like "cat," then it will also be doubtful

18 Blackburn (1984, 114–18) imposes the even stronger requirement that the speaker intend *all* of the intentions with which he utters e to be recognized, which is so strong as to exclude lying. Christensen 1997: 506 does the same, insisting that that all-inclusive intention is one of those that must be intended to be recognized.

19 As an alternative way of excluding such examples, Sperber and Wilson (1986a: 30–1, 60–1) require that S *intend to make it mutually manifest that he intends to make manifest the relevant belief*. Since this condition specifies only *what* is to be manifest, and says nothing about the *way* it is to be manifest, it fails to exclude examples like Stampe's.

20 Cf. Black 1972: 267; Holdcroft 1978: 145–8; Loar 1981: 213, 217, 247–53; Sperber & Wilson 1986a: 18–21, 31; Kemmerling 1986: 136–8; and Schiffer 1987a: 246. See Loar 1981: 247–53; Schiffer 1982: 124–5; and Grice 1982: 242 for replies.

that she means anything.[21] When the child gets a little older, it may be clear that her utterance indicates that she is thinking of a cat, without it being clear that she intends to indicate what she is thinking of. Until that is clear, though, "By 'cat' she meant 'cat'" and "She is expressing the idea of a cat" will seem to be overstatements.

A few have adopted the prohibition-clause strategy suggested by Grice, requiring the speaker *not* to have certain intentions, and thereby taking a negative approach to the explication of "openness." Holdcroft (1978: 144) requires that the speaker not intend any of his intentions to go unrecognized. Recanati (1986: 234) requires that the speaker have no intentions inconsistent with Gricean intentions of any order.[22] Neale (1992: 550) requires that the speaker not intend the audience to be deceived about his having the other intentions required for meaning. These requirements are too strong, as proven again by speakers who wish not to be identified or understood by their audience. Holdcroft's condition even excludes lying. The liar intends to deceive his audience, but does not intend that intention to be recognized. The infinitary conditions make it impossible to recognize speaker meaning.

My analysis also employs the prohibition-clause strategy: the speaker must not be covertly simulating an unintentional indication of the belief; hence he must *"undisguisedly" convey* the belief. Thus in food-tasting case (A) of §3.4, Stu did not mean anything by spitting out his food because he intended the action to be mistaken for an involuntary reaction to awful-tasting food. My requirement avoids the excesses of Holdcroft's and Recanati's by being more specific: the speaker must not be trying to conceal the *intentionality* of his belief indication in a certain way. There may be other aspects of his intentions that he is trying to conceal, such as the content of the belief that he is indicating, or the very identity of the subject of the intentions.[23]

The Stampe example cannot be excluded in exactly the same way, however. For Jack intended Jean to recognize that his smile was fake. Jack was overtly simulating, therefore, while in effect pretending to be covertly

21 For more on this point, see Bennett's (1976: 202–6) discussion of animal communication.

22 Cf. Platts 1979: III.3; Davies 1981: 11; 1987: 717; and Kemmerling 1986: 147.

23 Grice initially rejected the prohibition clause strategy because "it gave no explanation of *why* this was a reasonable condition to put into an account of speaker's meaning" (1982: 243). This is as misguided as rejecting a correct definition of "scalene" on the grounds that it provides no rationale as to why not having equal sides is a reasonable defining property. Grice was subsequently more receptive to the proposal, although he merely shifted the site of the alleged "ad hocness," and still thought that, "strictly speaking," meaning requires an infinite set of intentions.

simulating. We could exclude the example by modifying Theorem 3.5 (and Definition 3.6) to prohibit overt as well as covert simulation. But this would also exclude food-tasting case (B), which is a genuine case of meaning and expression. Alternatively, Theorem 3.5 could be strengthened by prohibiting simulations that are either covert *or semi-covert*, where "semi-covert" describes simulations designed to be both correctly taken as phony and incorrectly taken as designed to be taken as genuine.

While it would be easy to exclude examples like Stampe's, I believe that Theorem 3.5 is correct as it stands. Despite the number of distinguished scholars who have decided otherwise, I believe Jack *did* mean and express the belief that he had a good hand. I assume, of course, that even though Jack did not intend his smile to be mistaken for a genuine betrayal of emotion, he did intend it to be an indication that he thought he had a good hand. Given that, I believe that it is correct to say not only that Jack conveyed the idea and let Jean know that his hand was good, but also that he *gave a sign* and *signaled* that it was. His smile indicated that his hand was good, and in addition *he indicated* that it was. If, as Jack expected, Jean realizes that his phony smile is a sign that he has a good hand, then he even *communicated the information* to Jean. Indeed, if Jean was Jack's partner, others could justly complain that Jack had violated house rules prohibiting communication between partners. If someone else was Jack's partner, such as his wife, she could complain that he had thrown the hand by signaling to Jean. To be sure, Jack's action did not count as *telling* Jean about his hand. But meaning is known from other cases, such as signaling for a turn, to be broader than telling. Moreover, what he did was *tantamount* to telling her.

I believe there are at least two reasons why the literature has generally concluded that the speaker means nothing in cases of semi-covert simulation. First, in describing the cases, Griceans have focused on the speaker's intentions concerning what the audience is to believe and infer, which are inessential to meaning. Hence the speaker is never explicitly described as intending to provide an indication of his belief, which is essential. Second, Griceans have overemphasized paradigm cases. Indeed, Grice later claimed that to mean something is to be in the "optimal state with respect to communicating" (1982: 242). Cases with semi-covert simulation are clearly not paradigm cases of communication, and may not be optimal in some respect. But that does not prevent them from being cases of meaning, as verbal slips, bad diction, poor grammar, and a thousand other impediments to understanding prove conclusively. Bennett (1976: 126) put the argument quite baldly:

Whenever U contrives that he and A should be at cross-purposes in that way, the case seems not to be one of meaning, properly so called. Strawson explains why: meaning is essentially an attempt to communicate; but in real communication everything is open and above-board; and so meaning cannot exploit contrived cross-purposes.

Both premises are false. Meaning is not necessarily the attempt to communicate, as the dozens of cases canvassed in §4.1 and §4.2 show. Moreover, in real communication, everything is not necessarily open and above-board, as the ubiquity of lying and other forms of deception demonstrate. The hundreds of millions of dollars spent annually on advertising argue that it is an effective form of communication. To suggest that advertisers are always open and aboveboard would be a joke.

I grant that cases of meaning with semi-covert simulation are not as clear as others we have discussed. Since they rarely if ever occur in practice or thought, it should not be surprising that our intuitions about such cases are not the clearest. They are the proverbial borderline cases, and should be treated as such.

§4.4 COMMITMENT AND TRUTH CONDITIONS

Searle and Alston identified the basic defect of the Gricean approach as its attempt to define speaker meaning in terms of S's intention to perform a "perlocutionary" rather than an "illocutionary" act. Meaning involves not attempting to produce an effect in an audience, but rather trying to do something similar to (yet different from) asserting, ordering, asking, or promising. On my theory, the relevant illocutionary act is *expressing a belief*, that is, *providing an intentional indication that one believes a proposition*. For Vlach and Searle, it is *committing oneself to the truth of a proposition*. The idea that meaning involves commitment is the only significant alternative to the Gricean analysis hitherto presented.

What is commitment? For Vlach, it involves "attaching one's authority" to a proposition, which consists in asking an audience to "take one's word for" the proposition, and implies sanctioning the inference that it is true. Vlach's approach does not improve upon Grice's, however. A person who jokes, or talks to himself, or prefixes his remarks with "You shouldn't take my word for it" does not thereby commit himself (according to Vlach's definition) to the truth of what he says. And a person who signs a manifesto he has not read carefully may have unwittingly committed himself to the truth of proposition XXIII therein

even though he did not mean by signing it that that proposition is true.

A precursor of the commitment theory can be found in Alston, who held that to express a belief is to utter a sentence (or sentence surrogate) while recognizing that one's utterance is governed by a rule requiring one to have the belief. One thereby "takes responsibility" for having the belief, and is prepared to accept that a complaint is appropriate if one does not. By a rule, Alston meant a conventional norm "in force in a community," rather than a mere de facto regularity. Alston later indicated that "committing oneself to having a belief" and "lending one's authority to it" involve taking responsibility for having it.

Cases of presupposition show that Alston's definition of expression is too weak. One who says "My belief is strong" would typically be taking responsibility for having the belief that "my belief" refers to, but would not thereby be expressing that belief. Figurative speech, codes, verbal slips, and other deviations from convention show that Alston's analysis is too strong. One could express the belief that a movie is boring by saying "Even the ants went to sleep." Yet no rule requires one who utters this sentence to believe that the movie is boring. Codes have rules, but they need not be "in force" in any community. Finally, S may express the belief that H got her hair cut too short by making lawn mower motions and sounds, even though there are no rules of any kind connecting his gestures to the belief that he is expressing. The same goes for slips of the tongue.

Ambiguity presents a further dilemma. If there is any rule requiring one who utters "That is a bat" to believe that the indicated object is a baseball bat, then there is also a rule requiring one to believe that it is a mammalian bat. Yet one will typically not express both beliefs on any given occasion. Alston later claims that "There must be something either in the previous conversation or in the non-linguistic environment to single out" which belief one is taking responsibility for having (1965: 144). But any reasonably large sample of student papers will refute this contention. Suppose I ask you to identify the objects I describe by shape, including one baseball bat and one vampire bat. You might say "That is a bat," meaning and expressing the belief that the second object I described is a baseball bat, even though nothing in the context gives me a clue as to which belief you are expressing.

The case of ambiguity provides a vivid illustration of another important point. In general, *which belief S is required to possess, if any, depends on which belief S expresses*. So Alston got the order of determination backward.

When I utter "Steffi killed Martina," I would take responsibility for believing that Steffi beat Martina badly because I expressed that belief. I would not express that belief because I took responsibility for having it. Moreover, the rule according to which S must have the beliefs he expresses is the rule requiring us to be sincere and honest. That is a moral or prudential rule, however, not a linguistic rule or any other sort of convention. When morality does not require honesty, S can easily express a belief without recognizing any rule requiring him to have the belief. If a captured soldier under interrogation expresses the belief that an attack is planned for 10:00 A.M. in order to conceal the true time of the attack, he need not be taking responsibility for having that belief, and may regard anyone who complains that he lied as a fool or a war criminal.

A later attempt to cash out the notion of commitment can be found in Searle (1983: Chapter 6; 1986). Meaning that p presumably belongs to the "assertive" class of illocutionary acts, whose "defining point or purpose" is "to commit the speaker (in varying degrees) to the truth of the expressed proposition" (1986: 218). This is a matter of uttering something "with the intention that it represent a certain state of affairs," which is "at least partly the intention that certain conditions are conditions of success of the utterance," including the condition "that there must exist the relevant state of affairs." Since it is doubtful that many speakers have Searle's concept of a "success condition," there probably aren't many speakers who have the required intention. So it would be better to require simply that *S intends his utterance to be true iff p*. This accords nicely with Searle's claim to have ended the "Homeric struggle" by combining the truth-conditionalist and intentionalist theories of meaning.

Searle's analysis is incomplete, since he did not claim sufficiency for the indicated condition or any other. More seriously, the condition fails radically when the "utterance" is nonlinguistic. By raising my arm, I may mean that the enemy is retreating (Searle 1983: 167–8). Yet I, for one, believe firmly that the act of raising an arm cannot be described as true or false. Hence it is impossible for me to intend my arm raising to be true under any conditions. Searle's (1983: 164) claim that the mind "imposes conditions of satisfaction" on physical phenomena seems completely misguided.

Finally, Searle's proposal is implicitly, though perhaps not obviously, circular. Not having any intrinsic content, the truth value of a *sentence*, as opposed to a statement or belief, is relative to an interpretation. "The Sun is setting" may be true *in English* iff the Sun is setting, while being true *in*

S's code iff agent Smith's chances of success look good. "He flew to the bank" *as S meant it* may be true iff John moved quickly to the commercial bank, while being true *as T meant it* iff George moved by air to the river bank. It is natural, of course, to relativize the truth or falsity of *S's utterance* of a sentence to *S's interpretation* of the sentence. To make Searle's proposal work, "S's interpretation" must denote what S *meant*. And Searle will have to specify whether "S's interpretation" denotes what S meant *cognitively* or *cogitatively*, as a consideration of metaphor will show. So this path leads to circularity. Searle might try relativizing the truth value of S's utterance to *the contextually appropriate interpretation*. But as I understand this notion, S may have meant something contextually inappropriate. And if speaker meaning is one of the factors determining the contextually appropriate interpretation, circularity would again result.

Peacocke did not claim to be analyzing expression or meaning in terms of commitment, but his account can be seen as a variation on Searle's.

> Someone's utterance of a sentence s is an expression of his belief that p iff: s is intentionally produced (under its structural description), in part for the reason that he thinks that s is true, and he meets (I) [which specifies that the conditions whose obtaining would give him reason to judge that p are precisely those which give him reason to think that an utterance of s would in fact be true]. (Peacocke 1986: 115)

Peacocke's first condition is clearly not sufficient. If Fred is asked to utter some true sentences, he could comply by reading sentences from an almanac written in a language he does not understand, or by repeating sentences that his language instructor tells him are true. If Σ is one of the Italian sentences Fred has thus uncomprehendingly produced, it will not be true that Fred has expressed his belief that planets have elliptical orbits even if that is what Σ means. Peacocke's second condition secures this negative result. While Fred may be uttering Σ because he believes that it is true, his reasons for uttering it do not include his reasons for believing that planets have elliptical orbits.

Peacocke's condition I is too weak in one respect. Fred's reasons for thinking that an utterance of Σ would be true were precisely his reasons for judging that *my Italian teacher said Σ is true.* Yet by uttering Σ, Fred was not expressing his belief that his Italian teacher said Σ is true. Indeed, Peacocke's theory faces many of the difficulties of the Davidsonian analysis of word meaning, to which the resemblance is not entirely accidental. Since Sue is a pretty fair mathematician, when she utters "That polygon is three-angled" her reasons for thinking that her utterance is true are

precisely her reasons for thinking that the polygon is three-sided. Yet by "That polygon is three-*angled*," Sue need not *mean* that it is three-*sided*.

Condition I is problematic in another way. Suppose that Steve uses "It is raining" to express his belief that it is raining. Peacocke's first condition is clearly satisfied: Steve intentionally uttered "It is raining" in part because he believes that it is true. It is not so clear that Peacocke's second condition is met. While Steve's reasons for believing that sentence to be true will surely *include* his reasons for believing that it is raining, his reasons for believing the sentence to be true are *more extensive*. They also include facts about the meaning of the sentence, which are not among his reasons for believing that it is raining. Furthermore, we observed in §3.3 that Steve can use "I believe it is raining" to express not only the belief that he believes it is raining, but also the belief that it is raining. Peacocke's account does not allow this, since Steve's reasons for believing that it is raining are not his reasons for believing that "I believe it is raining" are true. Since Peacocke was only sketching a preliminary theory, he deliberately excluded from consideration sentences of the form "I believe that p." But it is hard to see how his theory could be generalized to cover them.

Cases of lying, fiction, metaphor, ambiguity, and nonlinguistic expression also show that Peacocke's general strategy of focusing on the speaker's reasons for believing the sentence he uttered to be true is even less likely to succeed than Searle's. All of Peacocke's problems can be avoided by focusing on what mental states the speaker intends his utterance or other action to be an indication of.

§4.5 COGITATIVE SPEAKER MEANING

Cogitative speaker meaning has been largely ignored, so there are few analyses to examine. Skulsky (1986: 593) offered one that purports to hold without regard to the cognitive/cogitative distinction.

> . . . "mean" in ["I mean y by x"] is short for "mean to *say*"; where y is a denotation rather than an expression, one *says y* by uttering an expression that *means* y. It seems to me that the most plausible unpacking of "I mean y by x" is "In using an expression x I mean (intend) to use an expression that means (denotes) y."

This is wide of the mark in the case of cognitive meaning. In our by now well-worn example, the novelist who writes "New York was nuked" intended to use an expression meaning "New York was attacked with nuclear weapons." But he did not mean *that* New York was so attacked.

In a rather different case, someone who says "The statement that God exists is false" used "God exists" in order to use an expression meaning "God exists." But the speaker did not mean by anything he said that God exists.

Conversely, if I say "Bill is a sea elephant" metaphorically, I may mean that he is obese. I know, though, that "Bill is a sea elephant" does not mean "Bill is obese." Skulsky's (1986: 593–4) theory is that while "Bill is a sea elephant" does not mean "Bill is obese" in "the vernacular" (English), it does mean that in the "impromptu dialect" I am introducing, which he calls "metaphorese." While there certainly is a (nonnatural) language that is otherwise like English except for the fact that "Bill is a sea elephant" means "Bill is obese," it is extremely unlikely that any normal speaker is even thinking about such a language when using a metaphor, much less using it (see Chapter 7). Moreover, if I am using a code in which "Bill is a sea elephant" means "Bill is obese," then I am not speaking metaphorically (cf. §2.4).

Even as an analysis of cogitative speaker meaning, Skulsky's suggestion has problems. Someone may mean something without intending to *say* anything. Thus I might put my fingers to my lips meaning "Be quiet!" precisely because I intend to say nothing and want others to say nothing. Even when speaking, S may mean something by a word that means something different or nothing at all. Suppose that S says "He is a farging ice hole" to avoid using curse words. It is quite clear what S meant by "farging" even though, as he realizes, "farging" means nothing. It is equally clear that S meant something unprintable by "ice hole" even though "ice hole" has only the most innocuous meanings. Finally, suppose I set myself the following literary task: to describe Jane as sexy while using a word that has as one of its meanings "shrewish woman." I can accomplish my goal by saying "Jane is a vixen." Then by using "vixen" I intended to use an expression that means "shrewish woman." But by "vixen" I meant "sexy woman," not "shrewish woman."

Schiffer (1972: 3, 110–11) had previously offered a similar analysis of cogitative meaning, with a significant improvement.

(3) S means μ by e iff μ is the sense or meaning of e S intends e to have (or be operative) on the occasion of his producing (or doing) e.

Schiffer's analysis avoids some of Skulsky's problems. By using "vixen" in the case just cited, I intended to use an expression meaning "shrewish woman." But I did not intend it to mean "shrewish woman" *on that occasion*. But Schiffer's analysis has the same trouble that Skulsky had with

the "farging ice hole" case. The speaker may mean something unprintable by this, even though he used this nonsense phrase deliberately to avoid using words with offensive meanings. The speaker does not believe that the phrase has that unprintable meaning, because he knows English, and does not want it to have that meaning, because then he would be swearing.

Schiffer's analysis fares even worse for the inclusive sense of cogitative speaker meaning. On that interpretation, the left-right conditional fails in two cases, *ellipsis* and *anaphora*, involving different devices used for the purpose of abbreviation. I might say "He can't win," meaning "Bush cannot win the upcoming presidential election." Yet I do not believe that the sentence "He can't win" means "Bush cannot win the upcoming presidential election," even on this occasion. Similarly, I might say "The grass does too," meaning "The grass in my lawn needs watering too." But the former does not mean the latter.

The reference to an occasion also means that even if (3) were universally true, it could not be correct as an analysis of what cogitative speaker meaning is. For *what a word means on any given occasion is determined by what the speaker means by it* (see §7.9). When S wrote "I destroyed the bank," the word "bank" meant "banking institution" there rather than "row of keys on an organ" in virtue of the fact that S meant "banking institution" rather than "row of organ keys." The fact that S meant "banking institution" is what makes it true that "bank" means "banking institution" on that occasion. If we define cogitative meaning in terms of applied meaning and applied meaning in terms of cogitative, our system of definitions is circular. Hence one of them must be incorrect as an analysis. The definitions I am offering have no circularity. Schiffer will have a hard time explaining what it is for a word to have a meaning on a particular occasion without referring to what the speaker meant by it on that occasion. He cannot say that e means μ on o iff (i) e means μ, and (ii) S intends e to mean μ on o; for that would be circular.

Grice (1969: 148–50) used compound sentences to distinguish cogitative from cognitive speaker meaning (in the inclusive sense). Suppose S uttered "Mary went if Jack did," meaning "Jack went" by "Jack did." S did not mean by uttering these words, however, *that* Jack went. My view has no trouble with parts of compound sentences: S used "Jack did" to express the thought, but not the belief, that Jack went. And he did that in order to express the compound thought (and perhaps the belief) that Jack went if Mary went. Grice hoped to define cogitative speaker meaning in terms of cognitive, as follows:

(4) S means "p" by e iff by some expression c(e) containing e as a proper or improper part, S means that c(p/e), where c(p/e) is the result of replacing e with p in c(e).

In the example just given, e is "Jack did"; p is "Jack went"; c(e) is "Mary went if Jack did"; and finally c(p/e) is "Mary went if Jack went." Grice's analysis thus nicely accommodates compound sentences.

Unfortunately, it is hard to see how Grice's analysis could be modified to account for the exclusive sense of cogitative speaker meaning – that in which S meant "Jack did," *not* "Jack went," by the words "Jack did" (§2.4). Moreover, Grice's theory assigns the wrong cogitative meaning in either sense in cases of irony. It rules, for example, that by the words "Rockefeller is in hock" S meant "Rockefeller is in great financial shape," since by saying "Rockefeller is in hock" S meant that Rockefeller is in great financial shape. It also rules that by "I cut myself" Steve meant "He cut himself," since he used the sentence to mean that he cut himself, not that I cut myself. In general, since Grice allows e to be an improper part of c(e) – as he must if isolated sentences are to have any cogitative meaning – his theory implies that the cogitative meaning of any complete sentence is the same as its cognitive meaning. Worse yet, Grice's theory assigns *no* cogitative meaning in fiction. For when a novelist writes "New York was nuked," his sentence has no cognitive speaker meaning at all. And no sentence, paragraph, or chapter containing the sentence has any cognitive meaning. Referring back to Grice's definition, p should be identical with e in this case; but there is no c(e) satisfying the right side of the equivalence. I see little hope that cogitative speaker meaning can be reduced to cognitive.

While almost as common as rain, fiction is bound to seem puzzling given the view that the primary function of declarative discourse is something like the expression or production of belief, and its corollary that cogitative speaker meaning is somehow derivative from or secondary to cognitive speaker meaning.

> From the point of view of the philosophy of language the problem of fiction is: how can the speaker utter a sentence with a certain meaning (whether literal or not) and yet not be committed to the truth conditions carried by that meaning? How for example does fictional discourse differ from lies? (Searle 1979: x)

The solution is perfectly trivial when expressing a thought is recognized as a speech act on a par with expressing a belief. Fiction is like nonfiction in the sense that both involve the expression of thoughts. But fiction is

unique in not involving the expression of beliefs, which is one reason why the novelist does not commit himself to the truth of what he says. Unlike the liar, a novelist who writes what he doesn't believe is not being insincere, since he is not expressing belief.

Very little attention has been paid to cogitative speaker meaning. So the inadequacy of extant analyses should hardly be surprising. What is remarkable is that despite the evident similarities between cogitative and cognitive speaker meaning, the two have been analyzed in radically different ways. I would submit that a condition of adequacy for any theory of speaker meaning is that the two types be analyzed similarly. Furthermore, I see no good reason for thinking that cognitive speaker meaning is more important, either theoretically or practically. To take one example, consider the attempt to reduce word meaning to speaker meaning, the more elusive part of the Gricean program. All attempts so far have used cognitive speaker meaning, or something like it, as the analysans. Yet the range of expressions by which we mean something cogitatively is much broader than that by which we mean something cognitively – including interrogatives, imperatives, and parts of sentences as well as whole declarative sentences. The greater directness of cogitative meaning also rules out a lot of nonliteral meanings. It would seem more natural, therefore, to pursue the reduction in terms of cogitative speaker meaning. That is what we shall do in Part II.

Both Schiffer (1982; 1987a: Chapter 9) and Bennett (1976) maintain that while their Gricean conditions may be too strong to cover all cases of speaker meaning, their conditions can nevertheless be used to analyze conventional word meaning.

> It would be futile to try to allow for the solitary cry of despair, for instance, in an absolutely general account of meaning which was to be a *basis* for an account of conventional meaning and thus of language. (Bennett 1976: 23–4)

This may have seemed futile to Bennett, but I shall try to show that it can be done in Part II. Schiffer and Bennett in effect claim that conventional word meaning can be understood independent of speaker meaning. I think this is doubtful. Surely what the words in a language mean is dependent at least in part on what speakers of the language mean by them. If no one had ever meant "female fox" by "vixen," "vixen" would not mean "female fox." Moreover, since Gricean intentions are inessential to speaker meaning, they should, a fortiori, be inessential to word meaning. Indeed, some of the most penetrating criticisms of the Gricean reduction of semantics

to psychology have been based on the specific sorts of intentions thought by Grice to be relevant. Thus Chomsky argued that communication is only one of the uses of language, which is more generally used for the expression of thoughts. The possibility is quite open, therefore, that the specific intentions mentioned in a correct account of speaker meaning will make it easier to account for word meaning.

5

Communication

Having defined speaker meaning in terms of expression and expression in terms of the intention to produce indications of belief or other mental states, we turn our attention now to communication and other concepts that entail communication. We have seen that the Gricean assumption that speaker meaning is the attempt to communicate distorted the Gricean analysis of speaker meaning in a number of ways, and we have suggested that his analysis would fit communicating, informing, and telling much better. We will now see that Grice's conditions are not completely appropriate even for these semantic acts. We will define communication in terms of expression and recognition, and then define informing and telling in terms of communicating. We will therefore adopt the Gricean strategy of reducing communication to intention, while again specifying different intentions from the Gricean ones. In the process, we will draw an important distinction between communicating *to* and communicating *with*, and reject the popular transmission model of communication.

Even though we have distinguished speaker meaning from communication, we will nonetheless go on to affirm in Part II that communication plays an important role in word meaning, as the common interest that sustains the conventions in virtue of which words have meaning. The use of words to express certain ideas is conventional, moreover, only when they are conventionally used to communicate them. So the idea that word meaning in living languages depends on communication in some way will be upheld.

According to Grice (1957), S meant that p iff S did something with the intention of producing in some audience A the belief that p by means of A's recognition of that intention. I argued in Chapter 4 that the *audience-oriented* character of the Gricean condition makes it completely unsuited for an analysis of speaker meaning, as does its emphasis on *recognition*. These very features are perfectly appropriate for *communication*, however. We cannot communicate unless we communicate something *to* someone. "S meant something," by contrast, does not require, or even allow, completion by "to ____." Hence a speaker cursing privately means something, but is obviously not communicating with anyone. Moreover, S cannot intend to communicate something to A without intending A to understand him. While A might understand the speaker's *words* without recognizing his intentions, A cannot understand the *speaker* without knowing what his intentions were in uttering those words. A speaker may mean something when talking to the dead, but that does not mean he is communicating with the dead or with anyone else. Grice's analysis of speaker meaning was driven in the wrong direction by his mistaken assumption that meaning is the attempt to communicate.

Communicating is not the same as attempting to communicate. People often attempt to communicate but fail. While Grice may have specified the wrong intentions, he was surely right in holding that speaker meaning is a matter of what the speaker intends. Intentions are not enough for communication, however. While *attempting* to communicate requires attempting to produce understanding, *success* in the former attempt requires success in the latter. That is, S cannot communicate with A unless A understands S. So in one important respect, Grice's condition is too weak to define communication.[1] An American tourist may very well have the usual meaning in mind when he says "I want a cup of coffee" in a Tibetan restaurant. But he will not communicate his desire unless the waiter understands him. Merely intending that the waiter understand does not suffice for communicating with him.

If attempting to produce a belief by means of recognition of intention defines the attempt to communicate, then producing it in that way defines communication. That is, the following analysis is implicit in Grice's assumption that meaning is the attempt to communicate.

1 So is that of Sperber and Wilson (1986a: 54, 63, 156), who require that A be *able* to recognize S's intentions, but not that A *does* recognize them. See also Recanati 1986: 214, 239.

(1) S communicates the belief that p to A iff S produces in A the belief that p by means of A's recognition of S's intention.

I will therefore refer to (1) as the *Gricean analysis of communication*. It is of course not Grice's analysis, since Grice never attempted to define communication.

While the shift to communication eliminates many problems, the Gricean analysis is still too strong. S can communicate the belief that p to A even though A's recognition that S intends to produce that very belief plays no role in its production, as when shouting "I am here" in a crowd. S can communicate a belief without producing it in any way, as when accusing his spouse of lying or proclaiming his innocence in the face of incriminating evidence. And S can communicate his belief while producing exactly the opposite belief in A, as in cases of countersuggestion and extreme mistrust. The Gricean analysis of communication is also too weak. If S knows that A mistakenly believes that "etymologist" means "entomologist," S can produce in A the (false) belief that Dr. Johnson is an entomologist by saying (truly) that Dr. Johnson is an etymologist. S might do this in order to see A make a complete fool of herself later. Nevertheless, the message that Dr. Johnson is an entomologist is not what S sent.[2] Since what S meant was not what A took him to mean, there was a failure of communication. If S had meant and communicated the belief that Dr. Johnson is an entomologist, S could be accused of lying. Yet he wasn't. He misled A without lying.

For S to communicate the belief that p to A, A must recognize that S means or implies p. That is, A must understand S. Whether A goes on to believe p or anything else, and how A arrives at her belief if she does, are irrelevant to whether communication has occurred.[3] We will accordingly design our definition of communication so that 5.1 follows as a theorem.

5.1 **Theorem:** *S communicates the belief that p to A iff S means (or implies) that p and is understood by A.*

More explicitly, S communicated the belief that p to A provided (i) S said or did something e by which S meant (or implied) that p, and (ii) A recognized by observing e that S thereby meant (or implied) that p.

A little-remarked failure of the Gricean analysis should be noted in this connection. On (1), S's communicating the belief p does *not* entail A's

2 Cf. Bach & Harnish 1979: 23, 31.

3 Cf. Black 1972: 270–1; Schiffer 1972: 121; Fodor 1975: 103–8; Bach & Harnish 1979: 154; Martinich 1984a: 19; Searle 1986: 212–18; Recanati 1986: 239; and Forbes 1989: 469.

recognizing *that S meant p*. It only entails A's recognition that S produced e with the intention of producing in A the belief that p. This failure of the Gricean analysis of communication could be avoided only by defining meaning that p without a recognition clause, as intending to produce in A the belief that p. But such a definition of meaning would be woefully inadequate, as Grice realized. Paradoxically, the Gricean assumption that meaning is the attempt to communicate, which ties meaning too closely to communication in some respects, prevents the Gricean from securing one of the most intuitive connections between meaning and communicating.

Harman (1974) claims that this failure can be avoided by defining speaker meaning in terms of a peculiarly self-referential intention. "We can say that S's intention is to produce a certain response r in A at least partly by virtue of A's recognition of this very intention" (Harman 1974: 225). I find such intentions literally unintelligible. To have such an intention, I would have to intend the state of affairs (or proposition) expressed by "I will produce a certain response in A in virtue of A's recognizing that I intend this very state of affairs (the one expressed by this whole sentence)." I do not know what it would be to intend such a thing, or to believe or desire the proposition expressed by that sentence. I cannot even get myself to think the thought expressed by that sentence. Since "the state of affairs expressed by this sentence" is the only description of the state of affairs provided by the sentence, its sense does not determine a referent. Consequently, I do not recognize myself as having such intentions, and cannot attribute them to others. I have never heard anyone expressing such intentions, although I have heard scholars in the grip of Gricean theory talk about them. I therefore do not think it can plausibly be insisted that everyone, including me, must have such intentions whenever they mean something.[4]

§5.2 COMMUNICATING *WITH*

According to Theorem 5.1, S can communicate something to A without expecting to. Hence S and A need not have any of the "mutual" knowledge or belief necessary for S to be assured of communicating.[5] Suppose, for example, that Sam is a pilot flying in formation with Ann. They are about to land, and their radios are out. Sam notices that only one of Ann's

4 See also Christensen 1997: 502–6, 517–20.

5 Cf. Blakemore 1992: 19. Contrast Bach & Harnish 1979: 5, 88 and Davidson 1983: 277. And see §9.8 of this volume.

wheels is down. In an attempt to alert her, Sam frantically points underneath Ann's plane, makes little circles with his hands, sticks up one finger, and so on. He hopes that his message is getting across, but worries that Ann might not be able to see what he is doing, let alone figure out what he means. Sam is trying to communicate, but does not expect to succeed. If Ann does understand, and puts down her other wheels, then Sam has communicated, despite his pessimism.

More radically, Theorem 5.1 allows that S may communicate something to A without even trying to, indeed, while trying not to. Consequently, A need not recognize, or even believe, that S intends to communicate anything to her.[6] Suppose that the German high command is ready to broadcast the time of the attack to field officers. To maintain secrecy, the radio operator sets the dial to a secure frequency, and turns on the scrambler. Unfortunately for the Germans, both the dial and the scrambler are inoperative. As a result, the Germans inadvertently broadcast the time of the attack to British agents, who are listening attentively. Similarly, if less momentously, the tennis player who insults the umpire in Rumanian to avoid a penalty may nevertheless communicate his feelings all too effectively if the umpire happens to understand Rumanian. Finally, Lawrence publishes a financial newsletter with a large circulation. He writes in one issue that IBM common stock is undervalued. One of his many readers is Edward, who rushes to purchase 5,000 shares. Then Lawrence has communicated his evaluation of IBM to Edward. But Lawrence was not trying to communicate anything to Edward in particular. Lawrence does not keep track of his subscribers, and may not even know that Edward exists. Alternatively, we might imagine that Edward is one of Lawrence's archrivals, and that Lawrence does his best not to communicate any valuable information to Edward.

This feature of Theorem 5.1 may appear to be implausible, because it seems clear, for example, that the Germans cannot be described as having *communicated with* the British. Nevertheless, I believe that it is correct to say that the Germans did *communicate the time to* the British. Similarly, the Rumanian communicated his feelings to the umpire but was not communicating with him. The proper conclusion to draw, I believe, is that the prepositions mark a conceptual difference. Communicating *with* A involves more than merely communicating something *to* A.

6 The contrary has been maintained in Bach & Harnish 1979: xv–xvi; Taylor 1980: 299; McDowell 1980: 130; Martinich 1984a: 19; Davidson 1986: 169; Recanati 1987a: 32; and Christensen 1997: 507. Compare and contrast Gilbert 1996: 251.

Griceans have emphasized the "*overtness*" of communication.[7] The requirement that S intend the audience to recognize his intentions was designed to express part of this feature. While almost as inessential for communicating something *to* an audience as for meaning, overtness does seem required for communicating *with* an audience. The examples just given suggest, in fact, that communication *with* A requires *intentionally* communicating something *to* A.

5.2 **Definition:** *S communicates with A iff S intentionally communicates something to A.*

Given Theorem 5.1, this would imply that communicating with A entails intending A to recognize what one means, which is partly a matter of what one intends. Note especially that since communicating with A is stronger than communicating something to A, the suggestion that the former requires intending to do the latter does not produce a circular analysis, entail an infinite regress of intentions, or make the intention to communicate in any way "reflexive."

As explained in §6.4, the use of individual rather than substitutional variables in Definition 5.2 signals that the contexts occupied by "S" and "A" are to be given a transparent interpretation, even though psychological predicates like "intentionally communicates something to Φ" will normally be given an opaque interpretation. Thus suppose that General Rumstadt told his most trusted aide, Fritz, the time of the attack. Unbeknownst to the general, Fritz was a mole passing secrets to the British. Then "General Rumstadt communicated with the mole" is unequivocally true. But while "Rumstadt intentionally communicated the time to Fritz" is true on both interpretations, "Rumstadt intentionally communicated the time to the mole" is true only on the transparent interpretation. Definition 5.2 says that a man communicates with a woman only if he communicates something to her under *some* description, but not necessarily under the description that we use to refer to her. General Rumstadt communicated with the mole because he intentionally communicated something to Fritz, who is the mole. The definiens of Definition 5.2 is not satisfied in the newsletter case. There is

7 Cf. Strawson 1964: 446–7, 454, 460; Bach & Harnish 1979: 5–8, 84–5, 152–4; McDowell 1980: 128–30; Blackburn 1984: 114–8; Martinich 1984a: 115–22; Sperber & Wilson 1986a: 30–1, 60–1; Kemmerling 1986: 147; Recanati 1986: 226–34, 238–9; Bertolet 1987: 205; and Avramides 1989: 50–5. See also §4.3 of this volume.

no description of Edward under which Lawrence does anything to him intentionally. Lawrence does intend to communicate certain things to his readers, but Edward is not identical to Lawrence's readers.[8] Edward might be the 359th reader in Lawrence's subscription list. But Lawrence need not intend to communicate anything to his 359th reader.

"To" and "with" mark a contrast in other contexts. Although people occasionally talk *to* the recently departed, talking *with* them is impossible. The distinction conveyed here is slightly different, however. S talks with A only if A talks with S. There must be a two-way conversation, with one person responding to the other. Both types of communication, though, can be one-way. S can communicate with A without A's communicating with S. This happens when the operator of a malfunctioning two-way radio can send but not receive.

It might be necessary to add another requirement for communicating with A. Suppose that I call Clay from home, with my wife Kathy beside me. I say, "Clay, do you want to play tennis? Great. I'll meet you at 3:00." I know that Kathy will overhear, and I want her to. Indeed, I said "I'll meet you at 3:00" to let Kathy as well as Clay know when I will meet him. I communicated the meeting time to both Kathy and Clay, and did so on purpose. Definition 5.2 rules, then, that I was communicating with Kathy as well as with Clay. I believe this is correct, but it is not as clear to me that I was communicating with Kathy as it is that I was communicating with Clay. If we wanted to distinguish between these cases, we could observe that I was *addressing* Clay but not Kathy. Communicating with A typically involves addressing A. The addressee is the person S is referring to when he uses the second person pronoun "you," and is therefore the person other than himself S is referring to when he uses the first person plural "we" to mean "you and I."[9] In nonverbal communication, the addressee is the person S makes eye contact with, taps on the shoulder,

8 Steve Kuhn has observed that further work is needed when A is a group or organization rather than an individual. Definition 5.2 correctly entails that if S is *communicating with his audience*, and his audience is the Audubon Society, then S is *communicating with the Audubon Society*. But the definition does not entail that if S is communicating with the Audubon Society, then S is *communicating with some member of the Society*. Yet that does seem to follow. Put another way, Definition 5.2 seems to correctly define "communicates with NP" when NP is a name or definite description, but not when it is a quantifier phrase like "some member of the Audubon Society."

9 An addressee requirement might therefore capture what Taylor (1980) had in mind when he said that communication requires making an object *"entre nous,"* which is supposed to be more than its being an object severally for me and you.

or singles out by a similar action. The addressee is the person S is talking *to*, or would be if he were to reexpress himself in words. In general, the speaker's meaning determines the addressee.

The addressee condition typically fails when the intentionality condition does. The German high command obviously did not address the British. And while Lawrence may have addressed his readers, he did not address Edward, even though Edward is one of his readers. The Rumanian example shows, however, that the two conditions are independent, and that the addressee condition alone is not sufficient to distinguish communicating with from communicating to. For the Rumanian did address the umpire. He failed to communicate with the umpire only because he did not intend the umpire to understand him.

While we could add an addressee clause, I am inclined to believe that Definition 5.2 is correct as it stands. If someone asked me whether or not I had communicated with Kathy at all today, I could not really say "No." I believe I did communicate with her obliquely about when I was going to play.

§5.3 TELLING AND INFORMING

The cases that show that communicating with A is stronger than communicating something to A also show that *telling* and *informing* A that p are stronger than communicating the belief that p to A. The Germans did not tell the British when the attack would begin, and the Rumanian did not inform the umpire that he was blind. Furthermore, an addressee clause seems mandatory for telling and informing. I told Clay that I would meet him at 3:00, and thereby informed him. I let Kathy know, but the process by which I communicated with her was too oblique to count as telling or informing her. This suggests that S told or informed A that p only if S intentionally communicated the belief that p to A by addressing A. It would follow from Definition 5.2 that telling A something, or informing A, entails communicating with A.

Telling differs from both communicating and informing in requiring the use of a *language, code, or signal system*. Telling entails *saying something* in the generic sense that is not restricted to speaking. Suppose that you ask S who won the contest. S makes eye contact and points at A, meaning that A won. A understands, and shrieks joyously. Since S succeeded in getting the good news across, he communicated with and informed A. Yet because S did not *say* anything, he did not *tell* A that she

won. Another requirement for telling is that what S said is the same as what S meant. The speaker cannot have used metaphor, or made a verbal slip.

Informing does not require saying. We can inform someone that A won by pointing. We can inform someone that we are going to turn by turning on a turn signal. Informing is stronger than telling, though, in requiring that what is communicated is knowledge. Informing differs markedly from misinforming. Someone who is lying, teasing, or ignorant can tell us something, but cannot inform us. "Johnny told his mother that he got an A, but he was lying" is an all-too-ordinary statement. Replace "told" with "informed" and the result is odd and contradictory. The fact that "S informed A that p" is sometimes used ironically when "p" is known to be false underscores the point.

In short, telling A something and informing him are two different ways of communicating with him. Whereas the concept of telling specifies the means by which S communicates with A, the concept of informing specifies the results of the communication. In both cases, S communicates a belief to A. In the case of informing, S must also communicate knowledge or information.[10]

§5.4 COMMUNICATING *TO*

In §5.1, we identified communicating the belief that p to A with meaning that p and being understood by A (Theorem 5.1). We then distinguished communicating *with* A from communicating something *to* A, and proceeded to define the former in terms of the latter. We also went on to analyze the stronger notions of telling and informing A that p in terms of communicating the belief that p to A. Let us now return to the basic concept in this group, that of communicating something to an audience. Theorem 5.1 is not a complete definition of communication, since it only deals with the communication of beliefs.

Besides our beliefs, we can communicate our thoughts, desires, intentions, hopes, fears, and emotions. We can communicate anything we can express. Suppose a novelist writes "New York was nuked," and thereby communicates to his readers the thought that New York was attacked with nuclear weapons. Theorem 5.1 simply does not apply. The novelist was not communicating the belief that New York was attacked with nuclear

10 See Davis 1999 for a complete analysis of informing and telling.

weapons, and did not mean by writing what he did that it was.[11] Insofar as he was writing pure fiction, the novelist did not mean that anything is the case. He did, however, *express the thought* that New York was nuked, as his readers recognized. Or suppose S spontaneously communicates his happiness to A by yelling "Yahoo!" Again, Theorem 5.1 does not apply. S was communicating happiness, but not the belief that he is happy. By screaming "Yahoo!" S did not mean *that* he is happy, or that anything else is the case.[12] S nevertheless *expressed* happiness, as A recognized. As the *American Heritage Dictionary* puts it, communication is "*effective expression*." To communicate is to express oneself in such a way that one is understood. So Theorem 5.1 can be generalized as follows. Let "M" be an individual variable ranging over "pure" mental states, including displeasure and the belief that $2 + 2 = 4$, but not the knowledge that $2 + 2 = 4$ (which entails a fact about numbers as well as a belief) or John's belief that $2 + 2 = 4$ (which entails that John has the belief).

5.3 **Definition:** *S communicates M to A iff S does something by which S expresses M and from which A recognizes that S is expressing M.*

Mental states can be expressed directly or indirectly. S might *imply* that he cannot play tennis, expressing that belief by expressing the belief that he has to study. If A recognized S's implication, both beliefs would have been communicated. If in Theorem 2.3, "means" is taken in its inclusive sense (§2.3), the parenthetical "or implies" is unnecessary. But officially we've defined the exclusive sense (Definition 2.1).

An important detail of Definition 5.3 is that A must recognize what S expressed by observing the action by which S expressed it. Suppose Alan tells Betty that *he won*, while Betty tells Cindy that *Alan told her he won*. Then Alan expressed the belief that he won. Assuming that Cindy believes Betty, both recognized that Alan expressed this belief. Yet Alan communicated that belief to Betty, but not to Cindy. The difference lies in how Betty and Cindy recognized what belief Alan expressed. Betty recognized what Alan expressed by hearing him say "I won," let us say,

11 The novelist did mean "New York was attacked with nuclear weapons" by his words, but this is a different notion of speaker meaning from meaning by uttering them that New York was so attacked, and involves the expression of thought rather than of belief. See §2.2.

12 There are cases, of course, in which we express happiness *by* expressing the belief that we are happy, as when we say "I am happy." In these cases, we communicate happiness by communicating a belief. But even here, communicating happiness cannot be identified with communicating the belief that we are happy, as the asymmetry of the "by" relation indicates.

while Cindy recognized what Alan expressed by hearing Betty say "Alan told me he won."

There is considerable debate over the precise cognitive process by which A recognizes what S has expressed. Some leading ideas are that understanding consists in directly perceiving meaning, decoding a signal, processing by a language processor, or inferring intentions.[13] This issue need not be decided here. A must recognize what S has expressed by observing S's means of expression. But the way in which observing S's action leads to A's recognition is, I believe, irrelevant to whether S communicated M to A. It need only be stressed that recognizing what S's words express is neither necessary nor sufficient for recognizing what S expressed. It is possible, for one thing, to understand S without knowing what the words he uttered mean. If S and A are both staring in amazement at a St. Bernard gobbling down huge quantities of food, and S says "*Le chien* is hungry," A may understand S even though she does not know French, and is not even sure that S is not using a metaphor or misspeaking. It is even possible to understand S while misunderstanding the words he uttered. If S and A both mistakenly believe that "entomologist" means "etymologist," S might communicate to A the belief that Ned is an etymologist by saying "Ned is an entomologist." Conversely, A may completely understand the words S uttered while having little idea what S meant, as when A overhears S say "I did it to him on the bank" without knowing to whom S was referring by "him," what action S had in mind, or which sense of "bank" S intended.

Some have claimed that recognition of the intention to communicate is necessary and sufficient for its fulfillment.[14] Both parts of this claim are incorrect. First, A may recognize that S intends to communicate something to her without S's fulfilling his intention to communicate. This may happen if A figures out what S is going to say before S manages to say it. Suppose, for example, that S is a stutterer. He is asked whether he wants lobster or steak. S tries to answer, and gets caught in a stutter on the word "I." A knows that S loves steak and hates lobster, and so correctly infers

13 See Black 1972: 273–4; Fodor 1975: 103–23; Bach & Harnish 1979: xv, 13, 21, 84–6; McDowell 1980: 137; 1981; Blackburn 1984: 113; 1992; Martinich 1984a: 129; Sperber & Wilson 1986a: Chapter 1; Grice 1986: 84; Millikan 1987: 726; Schiffer 1987a: 198, 214, 261–2; Avramides 1989: 16–17; Schiffrin 1994: Chapter 11; Laurence 1996: 285–6, 292; and Lance & O'Leary-Hawthorne 1997: 255. See also §8.6.

14 Searle 1969: 43, 47 and Bach & Harnish 1979: 13. Recanati (1986: 214–15, 222) argues that recognition is sufficient but not necessary. See also Strawson 1964; Schiffer 1972; and Blakemore 1992: 34.

that S intends to communicate to her his desire for steak. Despite the recognition of his intention, S has not yet communicated his desire for steak. The same thing occurs if the president schedules a news conference to announce some official business, and a reporter figures out what the president intends to announce before the conference takes place.

Second, S may intend to communicate something to A and do so, without A's recognizing that S intends to communicate it to her. Suppose S puts a note in A's mailbox that is unsigned and addressed to B. As planned, A reads the note and understands it fully, while inferring from the letter's address that the author did not intend to communicate anything to her. Alternatively, imagine that Smith has written a letter to Adams using a special code that makes perfectly good sense when read as English but says something entirely different when decoded. Smith's note says that Jones is in Berlin when decoded, and that Jones is on vacation when interpreted as English. Smith addresses the letter to Adams, intending him to decode it. Adams does so, understanding Smith to be expressing the belief that Jones is in Berlin. However, Adams has forgotten that Smith knows that he knows the code. So Adams assumes that Smith intended someone else to decode the letter, concluding that Smith did not intend to communicate to him the belief that Jones is in Berlin. Despite this misconception, Smith did communicate to Adams the information that Jones is in Berlin.

If S's intention to communicate something to A is to be fulfilled, A must recognize what S *expressed*, which depends on what S intends. But S need not recognize S's intention specifically to communicate. The Smith–Adams case shows that not even communicating *with* A requires A's recognition of S's intention to communicate something to him. Smith did communicate with Adams, even though Adams mistakenly assumed that Smith did not intend to.

§5.5 THE TRANSMISSION MODEL

It is often suggested that to communicate something is to *transmit* it to someone else.[15] The transmission model certainly fits electronic

15 See, for example, *Webster's Encyclopedic Unabridged Dictionary*; Augustine, *On Christian Doctrine*: Chapter 2.2; Bentham 1843: 329; Katz 1966: 98, 103–4; Bennett 1971: 1–2; 1976: 127–71; Barwise & Perry 1983: 120; Martinich 1984a: 10; as well as Sperber and Wilson's (1986a: Chapter 1) exposition and critique of the "Code Model," Schiffrin's (1994: Chapter 11) discussion of three models of communication, and Tormey's (1971) discussion of the "Expression Theory" in art. Contrast Bach 1994b: 11. Compare and contrast C. Doerge (in preparation), who defends the transmission model by arguing that all communication implies communication of thoughts.

communication, in which a sender transmits an electronic signal to a receiver. The transmission idea also applies to the markedly different sense in which people communicate diseases, as well as that in which actions or other events communicate information. If S communicates a disease to A, A gets the disease from S. And the sound of the engine communicates to the mechanic the information that the valves need adjustment only if the sound carries that information and the mechanic gets the information from the sound.[16] The transmission idea even fits the communication of thoughts by speakers. When a speaker communicates the thought that it is a nice day, he has the thought and expresses it, typically in words. The hearer hears the words, "decodes" them, and winds up having the thought herself.

The transmission model loses all plausibility, however, when we examine the communication by speakers of mental states other than occurrent thought. Imagine a known liar telling you that your son has cheated. The liar communicated the belief that your son has cheated. Yet he did not have the belief to begin with (since he was lying), and you did not end up with it (since you knew he was lying). So transmission of a belief is not necessary for communication. It is also insufficient. My belief that it will rain may cause me to carry an umbrella, which causes you to believe that it will rain. It does not follow that I communicated to you the belief that it will rain. Transmission is similarly irrelevant to the communication of desires, intentions, or emotions. The special character of the case of thought is due to the impossibility of expressing a thought without thinking that thought. With some exceptions for indexical thoughts,[17] the same goes for recognizing that someone else is expressing a thought. It is very easy, by contrast, to express a belief, desire, or emotion (or recognize that another is doing so) without having that belief, desire, or emotion.

Some statements of the form "S communicated ____ to A" do presuppose that S has the mental state communicated. These are obtained by filling in the blank with phrases consisting of a personal pronoun followed by a noun denoting a mental state. For example, "S communicated *his belief* that p" differs from "S communicated *the belief* that p" in presupposing that S does believe p. Otherwise the truth conditions of the two are the same. In general: *S communicates his ϕ to A iff S communicates ϕ to A, provided S has ϕ*. "S expressed his belief" and "S expressed the belief" differ in the same way. There is no implication that A acquire ϕ.

16 Again, see Dretske 1981.

17 See my forthcoming *Nondescriptive Meaning and Reference.*

Sentences of the form "S communicated ____ to A" have a different presupposition when the blank is filled in by expressions of the form "the F that p" in which F is a factive noun like "knowledge," "information," or "fact" itself. "S communicated the knowledge that p" presupposes that S knows p, and "S communicated the fact that p" presupposes that it is a fact that p. Otherwise these forms appear to have the same truth conditions as "S communicated the belief that p." *S communicates the F that p to A iff S communicates the belief that p to A, provided the F that p obtains (or is possessed by S).*[18] We noted earlier that informing is a special case of communication in which the speaker transmits knowledge and information to the hearer. But the communication of information to a speaker does not result in informing the speaker unless the hearer accepts the information as true. It is thus possible to communicate the same information over and over again, whereas it is not possible to inform the same person over and over again (unless she is very forgetful).

Finally, "communicated" can be followed by noun phrases like "the time of the attack," "the location of the mine," and "the number of voters." These phrases of course denote things that could not possibly be transmitted from speaker to hearer in any literal sense. S communicated the time of the attack if he communicated the fact that the time of the attack is whatever it happens to be. In general, *S communicates f(a) to A iff S communicates the belief that f(a) is G to A, for some G such that f(a) is G.* In presenting Definition 5.3, we stipulated that "M" stands for "*pure*" mental state names, and not factives. While Definition 5.3 defines the basic concept of communication, supplementary schemas like those presented in the previous three paragraphs are needed to define all instances of the form "S communicates ____ to A."

Some have argued that young children provide counterexamples to Gricean analyses of communication or understanding.

> If the contemporary position is correct that belief-desire talk expresses a folk theory of the mind instead of expressing concepts and knowledge that originate as introspective epistemological "givens," is it really plausible that extremely young children represent to themselves that the speaker *intends* that they should *believe* that p in the process of understanding their very first sentences? . . . It seems unlikely that tiny children have the requisite concepts to make this sort of inference. But

18 McDowell (1980: 127) criticized Strawson for focusing on the communication of beliefs, suggesting that "[i]f communication is conceived, by contrast, as the sharing of knowledge, it will not seem an accidental fact about communication that [it] is potentially helpful to its recipient." Unfortunately, given that so many communicators are either ignorant or mendacious, communication is often unhelpful or positively harmful.

if it is possible for very young children to become informed by speech without using belief-desire theory, should it not be equally possible for adults? (Millikan 1987: 726)

First, I do not believe that our concepts of belief, desire, thought, and intention are purely theoretical. On the contrary, I think it is evident that we can know directly, by introspection, what we believe, desire, think, and intend. Furthermore, children acquire these concepts at a remarkably early age. It is not at all evident that children know what words or speakers mean before they can recognize propositional attitudes. Second, I have not asserted, and have no need to claim, that any of the definitions I have offered are analytically true. I have argued, for example, that *S meant that p iff S undisguisedly said e as an indication that he occurrently believes p*. But I have not claimed that the formulas flanking this "iff" are *synonymous*. If it were true that young children can know what people mean without knowing what beliefs they intend to indicate, that would at most show that the definition is not analytic. It could still be true, and true in virtue of the nature of speaker meaning. Children can recognize that a man is drinking water before they know that he is drinking H_2O.

Third, Millikan's phrase "become informed by speech" is vague in a critical respect. It is quite clear that children acquire information in some way as a result of hearing speech. If I say "Mommy is in the kitchen," and three-year old Johnny runs into the kitchen calling "Mommy," then I transmitted a thought to Johnny, and made him believe the proposition I asserted. It is not at all clear that I informed Johnny that Mommy is in the kitchen. That would seem to require a cognitive sophistication that Johnny does not yet possess. In general, compare: *informing A, communicating with A, communicating a thought to A, transmitting a thought to A*. If the analyses I have offered are correct, these semantic acts are arranged in order of the decreasing amount of cognitive sophistication required of A. Inversely, I believe that our readiness to say these things of a young child increases along the list. It is much clearer that we are transmitting thoughts to very young children than it is that we are informing them. It is not implausible to assume that a child's first linguistic activities are much more rudimentary than an adult's. At first, words may elicit thoughts and thoughts words by simple associative processes; only later, when the child is more sophisticated cognitively, and more self-reflective, may the child truly understand the speaker and express thoughts in words. So the facts about young children may even be compatible with the additional claim that the definitions I have offered are analytically true.

6

Reference

This chapter will briefly introduce the concept of *speaker reference,* distinguishing it from the markedly different notion of *word reference*.[1] When S said "The president is tall," he may have been referring to Al Gore even though the words he uttered referred to George Bush. This difference between word reference and speaker reference may have occurred because of a false belief about who is president, a false belief about what "president" means, or a verbal slip. Alternatively, the divergence may have been intentional, as when S is using a code, speaking loosely or figuratively, acquiescing in someone else's unconventional usage, or conveying his feelings about who should be president.

A constraint on any theory of word reference is satisfaction of what I call the "disquotation formula": if it refers at all, *"Φ" refers to Φ*. "Detroit" (the name) refers to Detroit (the city), "Napoleon" refers to Napoleon, and so on. Putnam and Horwich have suggested that a theory of word reference would consist of nothing but such tautologies.[2] A Davidsonian might say that a theory of word reference is a recursive theory entailing an instance of the disquotation formula as a theorem for each referring term in the language. Most would like a theory explaining *why* "Φ" refers to Φ. No such theory is remotely possible for speaker reference. We cannot even say that *S referred to Φ iff S used "Φ."* S might use "Napoleon" tongue in cheek to refer to a mental patient who only thinks he is Napoleon. Or S might use it as a code name for an invasion. Or S might use "Napoleon"

1 This distinction has been clearly drawn in Kripke 1977; Fitch 1987: 8–11; and Bertolet 1987. See also Linsky 1963 and Bach 1987a: 3–6, 39–40. Word reference is often called semantic or linguistic reference. Similar distinctions were drawn in §2.1 for meaning and in §3.1 for expression.

2 Putnam 1981: 52; Horwich 1998a: 114, 118, 120–3; 1998b.

to refer to another person with that name. The best we can say along disquotational lines is that speakers *normally* or *conventionally* refer to Φ when they use "Φ." As a first approximation, word reference may be characterized as conventional speaker reference. While this idea is the germ of an explanatory theory of word reference, it does not tell us much about what it is for a speaker to refer to something. And the characterization is only an approximation (Chapter 10).

I will present a simple definition of speaker reference in terms of expression. I believe that the definition is accurate, but a full defense would be more of a diversion than a help in this work.[3] One of the many ambiguities that makes it seem problematic is the familiar distinction between *opaque* and *transparent* interpretations of psychological predicates. It will be critical throughout this work to keep that distinction sharp, so we will take some time clarifying it. Our default interpretation will be the opaque. The phenomenon of *intentionality* displayed by speaker reference in contradistinction to word reference will also be of general importance.

§6.1 SPEAKER REFERENCE

Kripke (1977: 263) asserted that the distinction between speaker reference and word reference is a "special case" of Grice's distinction between speaker meaning and word meaning (§2.1). There is certainly a strong connection. Imagine that S slipped and said "New Jersey" instead of "New York." Then S may have meant that New York is larger than Pennsylvania even though the sentence S uttered does not. Speaker reference depends on the speaker's intentions in much the same way that speaker meaning does. Indeed, since S meant that New York is larger than Pennsylvania by what he said, it *follows* that he was referring to New York. As for the other half of the distinction, "New York" refers to New York because it means "New York." Word meaning is related to conventional speaker meaning in much the same way that word reference is related to conventional speaker reference.

We will use "Φ" as a placeholder for all terms that can grammatically fill in the blank in formulas like "S referred to ____" or "S is thinking of ____." The placeholder "Φ" thus stands for what we shall call *object nominals*: nouns, noun phrases, gerunds, and nominalizations that can appear as direct or indirect objects. It cannot be replaced by a sentence, verb, adverb, or adjective. This means that speaker reference is not a

3 See my *Nondescriptive Meaning and Reference* for the full defense.

species of the cognitive sort of speaker meaning that was Grice's main focus. For Grice was primarily concerned with meaning that p, where "p" stands for declarative sentences. When "means" has its cogitative sense, however, it can occur with noun phrases (see §2.4). Kripke's idea that speaker reference is a special case of speaker meaning, when applied to the cogitative sense of meaning covered by Definition 2.5, suggests the following.

(1) S refers to Φ iff S means "Φ" iff S directly expresses the idea of Φ.

Referring to Φ, however, does not entail *direct* expression.[4] Suppose that S called Al Gore "the president." If this resulted from an erroneous belief about who is president, rather than a linguistic mistake, then S did not mean "Al Gore" by the phrase "the president." For S expressed the idea of Al Gore *indirectly*, by expressing the idea of the president. The two ideas are very different: one can occur to us when the other does not, for example; and millions believe that Al Gore is not the president. But since S believes that Al Gore is the president, he can express the former idea by expressing the latter. The same disparity occurs when ideas are expressed indirectly in other ways. Suppose that S employs a figure of speech and says "My conscience told me to give back the money," meaning that his wife Jane told him. S referred to Jane when he said this, even though he meant "my sense of right and wrong" rather than "Jane" by the words "my conscience." Of course, reference does not exclude direct expression either. When I said "π is approximately 3.14," I meant "π" and referred to π. Since it allows indirect expression, speaker reference is more akin to the *inclusive* sense of cogitative meaning than to the exclusive sense we have focused on (see §2.4).

In another respect, speaker reference is more specific than both cogitative meaning and expression. Like telling, referring entails *saying something* in the broad sense that includes writing and signing.[5] S must use a language, code, or symbol system containing words or wordlike elements (see Chapter 7). Referring requires *verbal expression*, we shall say. Suppose a policeman points at John, indicating that John is to pull over. Then the officer meant John, but did not refer to John. The reason the policeman's act does not count as referring is not simply that he did not use a natural language. We can refer to people in Morse code, American Sign Language, private codes, and artificial languages. The officer failed to refer

4 Cf. McKinsey 1978: 177.

5 Cf. Searle 1979: 142–3; Bertolet 1987: 203; and §5.3 here.

because he did not use any system of signs with the appropriate syntactic structure. Similarly, S may indicate that Maria has arrived by singing a few bars from her favorite song. Then even though S has produced words, he has not said anything in the relevant sense. Hence S did not refer to Maria even though he expressed the idea of Maria.[6] These considerations yield an equivalence that I believe holds universally:

6.1 **Definition:** *S refers to Φ iff S verbally expresses the idea of Φ.*

The precise sense of "verbal" intended in Definition 6.1 warrants further analysis. But since our main interest is in meaning and expression in general, we will not pursue greater clarity here. The definition is also likely to provoke many objections, because reference is associated with predication in a way that expression is not. I will argue elsewhere that these objections are not valid.

When "Φ" is something other than a proper name, various ambiguities may make Definition 6.1 seem to be obviously false. For example, when "Φ" is a *quantifier*, we find the familiar *scope ambiguity*. For example, "something" is normally used not to specify the object of reference, but to say that there is an object of reference. That is, "S referred to something" is normally interpreted as meaning "There is something that S referred to" (in which the quantifier has wide scope) rather than "What S referred to was something" (in which it has narrow scope). Definition 6.1 holds in either case, as long as the definiens is taken in the parallel sense. "S expressed the idea of something" should be interpreted as meaning "There is an idea that S expressed" in the former case, or as "S expressed the idea *something*" in the latter. Definition 6.1 holds as long as "Φ" is interpreted in the same way in both the definiens and the definiendum.

There is a more problematic ambiguity when "Φ" is a *definite description*. Let us stipulate that when Sam said "The president is tall," Sam was referring to Al Gore rather than to President Bush. Now consider this question: was Sam also referring to the president? According to Definition 6.1, the answer depends on the answer to another question: did Sam express the idea of the president? When we add that Sam thinks that Al Gore rather than George Bush is the president, we may conclude that the idea of the president is different from the idea of either Al Gore or George Bush and was expressed by Sam. This conclusion holds, however, only on the *attributive* interpretation of "the president." It is also possible to

6 Bertolet 1987: 211–12 attributes this example to John L. Pollock.

interpret the phrase *referentially*.[7] Indeed, since *we* know that George Bush is the president, it would be most natural for us to use "Was Sam referring to the president?" to ask whether Sam was referring to George Bush. And the answer to that question is "No." I myself find it most natural to interpret "the president" referentially in "Was Sam referring to the president?" and attributively in "Did Sam express the idea of the president?" Hence I find it natural to answer the first question "No" and the second "Yes." When I do this, Definition 6.1 seems too broad. But obviously, the term replacing "Φ" must be interpreted the same way wherever it occurs in instances of the generalization of Definition 6.1. When this is done, Definition 6.1 is confirmed. If "the president" is taken attributively in both places, "Sam referred to the president" and "Sam expressed the idea of the president" are both true. If the same phrase is taken referentially in both places, both statements are false.

§6.2 THE OPAQUE-TRANSPARENT DISTINCTION

An ambiguity easy to mistake for the referential-attributive distinction is especially troublesome when "Φ" is an indefinite description. Imagine the following. Susan and her mother have a serious difference of opinion about Ted, the man Susan intends to marry. As her mother's friend Edna walks in, Susan screams, "He's the finest human being I've ever met!" and stomps out of the room. Edna asks with more than a little curiosity, "Who was she referring to?" Susan's mother answers archly, "A crook."

(2) Susan is referring to a crook.

Let Susan's mother be right that Ted is a crook. Does it follow that Susan was referring to a crook? That Susan expressed the idea of a crook? On the most natural interpretation of the latter question, the answer is "No." "The idea of a crook" normally denotes the idea expressed by the words "a crook," which is not an idea that Susan expressed when she said "he"

7 Concerning this dual interpretation, see Arnauld 1662: 64; Donnellan 1966; Grice 1969: 141; Peacocke 1975: 208; Kripke 1977; Searle 1979: Chapter 6; Martinich 1979a; 1984a: Chapter 10; Wettstein 1981; Salmon 1982; Fitch 1987: Chapter 1; Bach 1987a: Chapters 5–6; Recanati 1989a: 299–300; 1993: Chapter 15; Cormack and Kempson 1991: 548; Neale 1992: 537–41; Elugardo 1997; Bezuidenhout 1997b; Reimer 1998; Berg 1999. Note that I do not mean to imply that the phrase "the F" itself is ambiguous, with both a referential and an attributive sense; see §10.5. It should not be assumed that if descriptions have just one meaning, then it must either be quantificational (à la Russell) or referential (à la Wettstein). Against Recanati, I would note that Sam may have *meant*, but he did not *say*, that Gore is tall (cf. §2.3).

(see Definition 7.9). The idea that she expressed was the idea of Tom, whom, we may assume, she believes to be a paragon of virtue. But on the most natural interpretation of the former question, the answer is "Yes." For what Susan's mother clearly meant was "The man Susan is referring to is a crook," by which she meant "Susan was referring to a certain man; and that man is a crook." This conjunction is true. When both questions are given the same interpretation, however, the answer is the same, confirming Definition 6.1. For example, if "Susan expressed the idea of a crook" is taken to mean "Susan expressed the idea of a certain man, and that man is a crook," then it is true.

A context that can be filled (or position that can be occupied) by singular terms is said to be *transparent* or *opaque* depending on whether or not the substitutivity of identity holds. Thus "Jack met ____" is a transparent context, because if "Jack met A" and "A = B" are true, then so is "Jack met B." It is impossible for Jack to meet George Bush without meeting the president. By contrast, "Jack uttered '____'" is opaque. Jack may well utter "George Bush" without uttering "the president," even though George Bush is the president. We will say that "Susan referred to a crook" and "Susan expressed the idea of a crook" are true in the *transparent* sense, false in the *opaque* sense. For the position occupied by "a crook" is transparent on the one interpretation, opaque on the other. From the fact that S referred to and expressed the idea of Phosphorus, it follows that he referred to and expressed the idea of Hesperus only on the transparent interpretation. Letting "Φ" stand for a pronoun-free object nominal, "S expressed the idea of Φ" on its opaque interpretation has the same meaning as "S expressed the idea 'Φ,'" which has no transparent interpretation. The idea "Hesperus" is distinct from the idea "Phosphorus," even though Hesperus is Phosphorus. We will be focusing on the opaque interpretation of psychological descriptions generally, and – with certain exceptions to be noted in §6.4 – will intend that interpretation in all of our theoretical formulations. Indeed, given Definition 7.9, we have the following theorem: *S refers to Φ iff S verbally expresses the idea "Φ."* As a piece of ordinary English, however, this holds only on the opaque interpretation of "S refers to Φ."

On the opaque interpretation of "S refers to Φ," "Φ" is intended to represent how the *subject* S conceives of the object he is referring to. Hence "Susan is referring to a crook" is true on the opaque interpretation only if Susan thinks of Ted *as a crook*. Since she does not think of Ted as a crook, the sentence is false. On the transparent interpretation, "Φ" is intended to represent the true nature of the referent, and thus reflects how

the *speaker* using "S refers to Φ" conceives of Φ.[8] "Susan is referring to a crook" is true in the transparent sense only if Ted is a crook, which is what Susan's mother thinks of him as. The two senses are therefore logically *independent*. In the example we have been using, "S is referring to a crook" is true on the transparent interpretation, false on the opaque. To reverse the truth values, imagine the opposite situation in which Susan mistakenly believes that Ted is a crook. Then when Susan refers to Ted as a crook, "S is referring to a crook" will be true on the opaque interpretation, false on the transparent. Hence we cannot describe either interpretation as broader or narrower than the other.

The opaque-transparent ambiguity is typical for contexts within the scope of psychological verbs. Thus on the opaque interpretation, "Susan wants to marry a crook" implies that Susan wants her husband to be a crook, and is false. On the transparent interpretation, the sentence implies "The man in question (the one Susan wants to marry) is a crook," which is true. For a vivid example involving belief, imagine that the speaker is describing Oedipus on his wedding night.

(3) Oedipus believes that his mother is in bed with him.

On its most natural interpretation, sentence (3) is false. Oedipus has not yet learned that Jocasta, the woman he married, is his mother. Had he believed that his mother was in bed with him, he would have been mortified rather than delighted, and surely would have behaved differently. But (3) can also be interpreted differently. Just imagine that you and the speaker normally think of Jocasta as Oedipus's mother, and that you have just asked the speaker where Oedipus thinks she is. Treat (3) so that the following is a pretty good paraphrase: "Oedipus believes that a certain woman, who happens to be his mother, is in bed with him." Interpreted in this way, (3) is undoubtedly true, assuming that Oedipus is not unconscious. In other words, (3) is false on the opaque interpretation, true on the transparent interpretation. If we imagined a radical reversal of the Oedipus story in which the king intended to commit incest and got in bed with Jocasta in the mistaken belief that she was his mother,

8 This is not to say, as Crimmins (1992: 167–8) does, that in an opaque report, the speaker uses "the words that the agent would use" to express the relevant concept. Susan may conceive of Ted as a crook even though she would not use the word "crook" because she does not speak English, or just does not like the word "crook." "S refers to Φ" is true on the opaque interpretation only if "Φ" is associated with S's concept in *our* language, namely English. The person we are talking about, S, need not associate that concept with "Φ" or any other expression. Compare and contrast Stich 1986: 142.

then (3) would be true on the opaque interpretation and false on the transparent.

Following Quine, many have assumed that sentence (4) differs from (3) in having only a transparent interpretation.

(4) Oedipus believes of his mother that she is in bed with him.

While the transparent interpretation of (4) may be the most natural, sentences of this form *can* be interpreted opaquely, at least with a little effort.[9] Lois Lane believes of Superman, but not of Clark Kent, that he can fly.

Many other terms have been used to mark the opaque-transparent distinction, principally "*de dicto–de re*" and "notional-relational."[10] We will avoid these terms for at least two reasons. First, even on the transparent interpretation, "S refers to Φ" is not aptly described as *de re* because it is not relational. For a singular term position to be relational, existential generalization must be valid as well as the substitutivity of identity. But existential generalization is not valid for the Φ position in "S refers to Φ." "John is referring to a nonexistent person" may be true in the transparent sense because he is referring to Santa Claus, who does not exist (see §6.3). Similarly, John believes *of Santa Claus* that he brings presents at Christmas, so the transparent interpretation of belief sentences is not relational either. Second, sentences that would generally be cited as describing clear cases of *de re* beliefs can also be interpreted opaquely. We noted in connection with (4) that the "believes of Φ" locution can carry the opaque interpretation as well as the more familiar transparent interpretation. Consider

9 Cf. Bach 1987a: 198; Recanati 1993: 331.

10 Quine (1960: §30), Fodor (1981: 234), Williams (1991: 144), Bach (1987a: Chapter 10), Devitt (1989a: 87–8), Cormack and Kempson (1991: 566), and Recanati (1993: Chapter 17) use "opaque" and "transparent." Perry (1979: 7–8), Stich (1983: 34, 73, Chapter 6), Noonan (1984), Zalta (1989: 460, 469), Fine (1989: 254–5), Fodor (1990a: 73), Horwich (1998c: §2), and Moore (1999a: 345, 359–60) use "*de dicto*" and "*de re*." Quine (1956) and Evans (1982: 132) use "notional" and "relational." Russell (1905: 209) uses "secondary" and "primary." Hungerland (1960: 221) uses "intentional" and "extentional." Burge (1979a: 538) uses "oblique" and "nonoblique." Barwise and Perry (1983: 201–3, 259) use "inner" and "outer." Kapitan (1994: 274) uses "internal" and "external." Castañeda (1977: §1, §3) uses "propositionally transparent" and "propositionally opaque," and notes that a description is *referentially* opaque/transparent iff it is propositionally transparent/opaque. See also Loar 1972: 46–7; Sadock 1974: 46; McKinsey 1978: 179; 1994; Burge 1977; 1979b: 538; Brand 1983: 183; Lycan 1985: 90–92; Wettstein 1986: 205–8; Boër and Lycan 1986: 71, 142–4, 147–8; Bach 1987a: 17; Forbes 1987: 9; Richard 1989; Crimmins and Perry 1989: 708; D. F. Austin 1990: Appendix; Kiteley 1991: 373–5; Crimmins 1992: 33, 170ff.; 1995a: 465; 1998: 18–22; and Reimer 1995. Dennett (1982: 54–90) shows that *de re* belief is seriously indeterminate in cases of mistaken identity, and argues that the notion is predictively and explanatorily empty.

also Richard's (1989) phone booth example, in which I am talking to you over the phone while watching a woman across the street in a phone booth who, unbeknownst to me, is you. If I see a bulldozer bearing down on the woman in the phone booth, then "I believe she is in danger" will be true. But since I do not realize the woman I am talking to is the one in the phone booth, "I believe you are in danger" will be false. Since "you" and "she" refer to the same woman, these belief statements must have the opaque interpretation. But believing that *she* is in danger and believing that *you* are not would commonly be classified as *de re* beliefs.

Despite their independent truth conditions, the opaque and transparent senses of reference are closely related conceptually, with the opaque sense being primary. In general, a statement of what someone is referring to is true in the transparent sense in virtue of the fact that a statement of the same form is true in the opaque sense. That is, whenever "S refers to Φ" is intended in the transparent sense, it will always be possible to find or introduce a term "Φ′" making the following *transparent-to-opaque reduction formula* true:

(5) *"S referred to Φ" is true in the transparent sense because (i) Φ′ is Φ and (ii) "S refers to Φ′" is true in the opaque sense.*

Thus "Susan referred to a crook" is true in the transparent sense because "referred to Ted" is true in the opaque sense, and because Ted is a crook. In this example, "Φ" is "a crook" and "Φ′" is "Ted." Clause (i) ensures the substitutivity of identity. For if "A = B" is true, then "A is B" will be true. (When "Φ′" is a plural term like "cats," "is" must give way to "are" in (i).) Similar transparent-to-opaque reduction rules hold for "S expressed the idea of Φ" and for propositional attitudes. Thus (3) is true in the transparent sense because Jocasta is Oedipus's mother, and because "Oedipus believes that Jocasta is in bed with him" is true in the opaque sense.

It follows from the reduction formula that while the transparent sense is not broader or narrower than the opaque, the class of terms that correctly describes what S is referring to in the transparent sense on any given occasion is broader than the class that correctly describes what S is referring to in the opaque sense. Any term *coextensive* with "Φ′" will describe what S is referring to in the transparent sense, but only terms *synonymous* with "Φ′" will describe what S is referring to in the opaque sense (cf. §13.6).

"Latitudinarianism" is the view that "S referred to Φ" is true in the transparent sense iff for *any* term "Φ′," "S referred to Φ′" and "Φ′ = Φ"

are both true.[11] On this view, that is, "S referred to Φ" is true transparently provided that S referred to Φ′ for *some* Φ′ = Φ. Many have argued, on the contrary, that in order for the reduction formula to be true, "Φ′" must be demonstrative or at least rigid.[12] To update and adapt Perry's (1979: 620) example, suppose that S asserts tautologically, "The president is the president," using "the president" attributively to refer in the opaque sense to the president of the United States. It does not appear to follow that S referred to Bush in any sense, even though Bush is the president. Or suppose Bill says to me, "The chair of the Budget Committee, by contrast, does not know how to count" after praising my financial savvy, not realizing that I am the chair of that committee. Is "Bill was referring to me" true or not? This is not at all clear, even though (i) and (ii) of the reduction formula are satisfied when "Φ′" is "the chair of the Budget Committee" and "Φ" is "me." Sosa (1970: 894–5) argued that the class of appropriate terms is context-dependent. There do appear to be contexts in which referring to the president would make it true that one referred to Bush. Suppose the speaker answers "The president" when asked whether anyone knows the nuclear codes, without knowing who the president is. We would have to deal with this issue if we were going to use (5) to define the transparent interpretation in terms of the opaque, or if we were going to provide a serious generative semantics for the transparent sense. For then we would have to define the appropriate class of terms. But since we are generally going to focus on the opaque interpretations of psychological descriptions, we will not attempt this task (but note §6.4).

Formula (5) suggests that the opaque-transparent ambiguity is an *amphiboly*. The transparent sense appears to result from a deep structure in which "Φ" is outside the scope of the underlying verb, which is then transformed into a surface structure with "Φ" inside the scope of the verb. The opaque sense would result from a deep structure in which "Φ" is inside the scope of the underlying verb. If we represent the scope of the underlying verb with brackets, the deep structure of (2) on the opaque interpretation might be represented informally by "Susan referred to [a crook]," while the deep structure on the transparent interpretation might be "Susan referred to [someone] (who is a crook)." If this is correct, then the opaque-transparent ambiguity is related to the scope ambiguity

11 See Noonan 1984: 218 and E. Sosa 1995.

12 Cf. Kaplan 1969; 1989: 606; Sosa 1970: 887; Loar 1972: fn. 6; 1976a: 368; Donnellan 1977; Schiffer 1978: 173, 202–3; 1981: 50–1, 86; Lewis 1979a: 539; Chisholm 1976: 168–70; 1981: 117–18; Fitch 1987: 126; R. M. Adams 1989: 36–7; Crimmins 1992: 171–9; 1995a: 476; Feit 2000: 42.

found in "S referred to something." In parallel fashion, we may represent its two deep structures by "Something is such that S referred to [it]" (wide scope) and "S referred to [something]" (narrow scope). The transparent interpretation of "S referred to a crook" thus resembles the wide-scope interpretation of "S referred to something," and the opaque interpretation of the former resembles the narrow-scope interpretation of the latter.

Some have argued that the opaque/transparent distinction is not strictly speaking an *ambiguity*. Instead, they view intentional statements as *contextually variable* in some way that is more like *indexicality*.[13] There is some justice to this view. The terms "flat" and "straight," for example, can be interpreted according to variably strict standards. On a "strict" interpretation, they apply only to objects that are perfectly flat or straight. On a "loose" interpretation, they apply to objects that are sufficiently close to being straight or flat for the purposes of the discussion. Taken strictly, the terrain in Iowa is not flat, since there are some low hills. Taken loosely, Iowa is flat. Compared to Nepal, it is "flat as a pancake." Despite the fact that different interpretations are possible, we do not ordinarily classify "flat" as ambiguous.[14] We do not ordinarily classify it as indexical, either. Although we have to infer the appropriate standard from the context, the context never determines the extension of "flat" in the way that it determines the extension of "I," "that," or even "foreign." The opaque interpretation of intentional statements may be viewed as the "strict" interpretation, since substitution is permitted only if terms are synonymous. The transparent interpretation is "loose," since substitution is permitted even when the terms are not synonymous, as long as they are coextensive. The speaker's intentions rather than the context determine the interpretation. It would take us too far afield to try to settle this controversy here. I will continue to refer to the opaque/transparent distinction as an ambiguity. But nothing will hang on whether "ambiguity" is the correct term for it. What is important is that psychological descriptions generally have more than one interpretation, and it is essential not to confuse them. One and the same sentence can be used to "say" different things, and whether it is true or false depends on how it is taken.

13 See Stich 1983: Chapter 6; 1986; Richard 1989; Crimmins 1992; Recanati 1993: 391–2.

14 On Stich's view, belief sentences are similarity claims, which are true only relative to a standard of similarity, which varies from context to context. In a discussion of political systems, "Cuba is similar to the USSR" would have been counted true in 1983; in a discussion of climate, the same sentence would have been counted false. Yet these facts do not show that "similar" is strictly speaking ambiguous. Stich's similarity analysis of belief is, for a variety of reasons, quite wide of the mark. See §13.2.

Whether transparent or opaque descriptions are more useful depends on the context. Some purposes, including those that we are most interested in, require the opaque interpretation. *First*, only when psychological descriptions are interpreted opaquely can their truth value be recognized by the subject introspectively. The fact that Oedipus would strenuously deny (3) is therefore evidence against its truth only on the opaque interpretation. To know that (3) is true on the transparent interpretation, we not only have to know something about Oedipus's beliefs, we also have to know that Jocasta is Oedipus's mother. For similar reasons, S's assertions of not-p are compelling evidence against "S believes that p" only when that belief description is taken opaquely. Sentence (5) makes it clear that referring to Φ is a *purely* psychological act only on the opaque interpretation. For whether or not S refers to Φ in the transparent sense depends on whether or not, as a matter of objective fact, Φ' is Φ. It is neither necessary nor sufficient for S to realize in any way that Φ' is Φ.

Second, accusations of inconsistency require opaque descriptions of S's beliefs. For example, interpreted either way, the sentences in (6) should be true.

(6) Oedipus believes that people in bed with their mothers are committing incest.
Oedipus believes that he is not committing incest.

We could conclude that Oedipus has inconsistent beliefs only if (3) were true on its opaque interpretation. Since it is not, Oedipus's mistake is less serious from a logical or moral point of view.[15] Looked at in another way, given "S believes p" we can confidently predict "S does not believe not-p" only when these forms are taken opaquely. For whereas it is difficult or impossible to find instances where both are true opaquely, there are many instances in which both are true transparently. Thus when Stan points to one picture of Jones and says sincerely "He is the murderer," and then points to another picture of Jones and says equally sincerely "He is not the murderer," Stan believes of Jones both that he is and that he is not the murderer. Hence "Stan believes that Jones is the murderer" and "Stan believes that Jones is not the murderer" are both true in the transparent sense. It would be very remarkable if both were true opaquely.

15 "Mates [1958: 240] provides a nice example of an extensional use of 'believes'; 'That gullible juror actually believes that the murderer had nothing to do with the crime'"(Hungerland 1960: 221). The example is particularly clear because interpreted opaquely, the speaker would be saying that the juror believes a self-contradiction. Since the juror is surely not that gullible, charity leads us to select the transparent interpretation. See also Schiffer 1978: 178 and Dennett 1982: 87.

Similarly, S's learning that p implies a change in the truth-value of "S believes that p" from false to true only when the belief statement is taken opaquely. Interpreted transparently, for example, "Oedipus believes that he is married to his mother" was true both before and after he learned the awful truth.

Third, given the greater frequency with which "S believes p" and "S believes not-p" are both true transparently, it is more difficult to predict the subject's behavior on the basis of transparent belief descriptions. For example, interpreted transparently (or opaquely), Oedipus believes that Jocasta is his wife and wants to sleep with his wife; interpreted transparently (but not opaquely), Oedipus also believes that Jocasta is his mother and does not want to commit incest. Given what Oedipus believes and desires described transparently, we could not predict whether or not he will sleep with Jocasta. Clearly, what we should base our prediction on is what he believes and desires described opaquely. Correlatively, it is the fact that "Oedipus wants to sleep with his wife" is true in the opaque sense that explains his behavior, not the fact that he wants transparently to sleep with his wife. For that fact depends on objective facts about Jocasta that are not involved in the causation of his behavior, and that could be altered without any change in his behavior.

Finally, (5) makes it clear that we need to know what S is referring to in the opaque sense in order to determine and explain what S is referring to in the transparent sense.

There are other purposes, however, for which knowledge of transparent descriptions of psychological states is necessary or at least sufficient. Suppose that S says "The tayra is a primate." If A does not have the concept of a tayra, we will not be able to tell her what S is referring to in the opaque sense. A will not understand "S is referring to the tayra" or any synonymous sentence. Nevertheless, we can give A useful information by telling her that S is referring to a weasel, which holds in the transparent sense. This information will suffice, for example, for A to determine that what S said was false. Similarly, if S says "I want a book about the thirty-fifth president," and the bookstore clerk does not know who the thirty-fifth president was, I may be able to help by telling her that S was referring to Jackie Kennedy's first husband. For even if S had never thought of the thirty-fifth president as Jackie's first husband, the clerk will be in a position to satisfy S's desire if she knows who Jackie's first husband was. Finally, if I know that S is trying in the transparent sense to find a mythical creature, I know enough to conclude that S will fail and be disappointed or frustrated.

An individual's thoughts are not limited to those that can be expressed conventionally in any particular language. Consequently, a speaker may refer to an object for which there is no word in the language we are using to describe the speaker. Inventions and discoveries make this abundantly clear, as do radically different cultures. More unusual, but hardly rare, is the case in which there are words in our language with the right extension but none with the right intension. That is, the language that we are using may provide us with appropriate transparent descriptions of the speaker's referent, none of which are appropriate as opaque descriptions. A variant of Kripke's puzzle (1979) illustrates this.[16] Suppose that Peter believes that there are two famous Americans named "Ronald W. Reagan," one an actor, the other a president. When watching a movie, Peter says "Reagan is nothing but a two-bit actor." When watching a presidential address, he says "Reagan is one of the greatest Americans who has ever lived." It is easy to provide a transparent description of Peter's referent: he is referring to the same man in both cases, namely Ronald W. Reagan. But clearly, Peter is expressing different ideas on the two occasions. As a result, he is not contradicting himself. Yet there are no words in the English language (or in any idiolect that we share with Peter) that express those ideas. So we have no conventional way of providing an opaque description of what Peter is referring to. The best we can do is to improvise, saying that Peter is referring to Ronald Reagan *the actor* in the first case, and to Ronald Reagan *the president* in the second, noting that Peter thinks the actor and the president are different people. What we are doing cannot quite be described as "disambiguating" the term "Ronald Reagan." For it conventionally applies to the president and the actor in exactly the same sense. Nevertheless, the italicized phrases serve an analogous function.[17] Kripke's puzzle will be discussed further in §13.2.

According to Burge, demonstrative reference shows that what S refers to is not determined by the content of the thought S expressed.

16 See also Vendler 1976: 42; Pollock 1980: 489; Noonan 1980–1; Lewis 1981a; Linsky 1983: 144; and Loar 1987; 1988; as well as Perry 1977: 483; Evans 1982: 84; Barwise and Perry 1983: 252; Castañeda 1985; Salmon 1986: 120–1; Böer and Lycan 1986: §4.2; Kaplan 1989: 108–9; Crimmins and Perry 1989; Devitt 1989a: 91–2; Forbes 1990: 557–62; Taschek 1995a: 277–9; D. Sosa 1996; Frances 1998; and Moore 1999a.

17 Loar argues that "S believes that Reagan is a man" may be true in virtue of two *different* beliefs S has even when it is given a *univocal* reading. This phenomenon is ubiquitous with transparent descriptions, impossible with opaque descriptions. The problem in the case at hand is that we have no conventional opaque descriptions. Carey (1991) argues for the converse possibility in which a subject expresses one concept where we make a distinction.

When Alfred refers to an apple, saying to himself "That is wholesome," what he refers to depends not just on the content of what he says or thinks, but on what apple is before him. Without altering the meaning of Alfred's utterance, the nature of his perceptual experiences, or his physical acts or dispositions, we could conceive an exchange of the actual apple for another one that is indistinguishable to Alfred. We would thereby conceive him as referring to something different and even as saying something with a different truth value. (1979b: 544)[18]

It is true that Alfred is referring to a different apple when the apple is switched. But it is also true that Alfred is thinking about a different apple. The reason is that "the same apple" and "a different apple" are interpreted in both contexts as *transparent* descriptions. It is true for the same reason that S is thinking of the same Ronald Reagan, even though he thinks they are different people and is expressing different ideas. In general, comparative nouns like *"the same object"* and *"a different object"* are interpreted as transparent descriptions when they replace "Φ" in either the definiens or definiendum of Definition 6.1. What these examples show is not that reference is underdetermined by thought content, but rather that transparent descriptions of objects of both reference and thought are underdetermined by opaque descriptions. That is why the reduction formula has two clauses.

Failure to distinguish between the opaque and transparent senses of psychological descriptions can lead to serious conundrums. Here is a passage from a leading textbook used to introduce thousands of students to cognitive science.

> Propositional attitudes (henceforth, simply "beliefs") are problematic for a number of reasons. For our purposes, the central problem is that they look like the things that should be identified and grouped according to their content, but real scientific and commonsense difficulties seem to stand in the way of doing that.
>
> . . . Suppose that Kilroy is the world's leading cognitive scientist, a renowned goatherd, and Sam's next door neighbor. Betty knows nothing of Kilroy's glorious caprine successes but has long admired his work in cognitive science. Sam, on the other hand, though blissfully ignorant of Kilroy's career in cognitive science, is in awe of his champion goats. Now, suppose that as Sam and Betty are enjoying a beer at the local dive, in strolls Kilroy. Betty comes to believe *that the world's greatest cognitive scientist has entered the bar.* Sam, on the other hand, comes to believe *that the world's greatest goatherd has entered the bar.* Both are right, of course, and what makes them right is that they both correctly believe *that Kilroy has entered the bar.* The very same fact makes both of their beliefs true. Both of their beliefs are

18 See also Fodor 1981: 236–7, 331; Peacocke 1981: 198; Stich 1983: 65; Bach 1987a: 16.

about Kilroy. In fact, a mind-reading bartender would truthfully report that both of them believe *that Kilroy has entered the bar.* And there's the problem. On the one hand, if the content of their belief is what matters, they do seem to believe the same thing. On the other hand, if the functional role that belief plays in their internal information-processing system is what matters, they do not believe the same thing, since although Betty's belief is connected with other beliefs about cognitive science, Sam's are connected with other beliefs about goats. Hence, they have different relations to other beliefs, inputs, outputs, and so forth, with the result that on any functionalist account they are different beliefs. Therefore, if cognitive science is to realize the dream of a functionalist account of the mind as an information-processing system, and if it is to account for beliefs in this scheme – to treat beliefs, not as classified by their content, but as classified by their functional role in the internal economy of the information-processing system – the question then is, How can cognitive science do this, while at the same time doing justice to the obvious fact that beliefs are beliefs just because they are about things? (Stillings et al. 1995: 356)[19]

Sam and Betty "believe the same thing" only in the transparent sense. In the opaque sense they believe different things. The principle that beliefs should be identified and grouped according to their content does hold if (i) the content of a belief is defined as what the subject believes in the opaque sense, and (ii) we have certain interests, such as an interest in the causal role of beliefs. It is only in the opaque sense that we can say that Betty's belief is connected with other beliefs about cognitive science, whereas Sam's is connected with other beliefs about goats. We cannot really talk about the functional role of believing that Φ is P in the transparent sense. For as the analog of (5) for belief would make clear, S may believe that Φ is P in the transparent sense in virtue of having many different beliefs in the opaque sense, with very different functional roles.

We can also maintain that beliefs should be grouped according to their content if (iii) the content of a belief is defined as what the subject believes in the transparent sense, and (iv) our interests lie in the truth conditions of beliefs. Both ways of classifying beliefs are important, but for different purposes.

Given that our main purpose is understanding meaning, it will be most important for us to classify according to what subjects believe in the opaque sense. The same goes for what subjects refer to, think about, desire, and intend. The opaque interpretation will therefore be our default interpretation. Whenever possible, terms I use in intentional contexts

19 Contrast Böer and Lycan (1986: 153), who focus on indexical beliefs.

are to be interpreted as opaque, attributive descriptions, unless I specify otherwise.

§6.3 INTENTIONALITY

On our default interpretation, "S referred to Φ" and "S expressed the idea of Φ" are *opaque* in their "Φ" position. This means that the substitutivity of identity fails: "S referred to Φ" may be true and "S referred to Φ′" may be false, even though $\Phi = \Phi'$. Both contexts are also *intentional* in Brentano's sense: neither entails that Φ exists. That is, existential instantiation fails. When a five-year-old says happily, "Santa Claus is going to bring me a new Barbie doll," she is referring to Santa Claus, and expressing the idea of Santa Claus. While the idea of Santa Claus exists, Santa Claus does not. The child may similarly think about Santa Claus, want him to bring her a new Barbie doll, and expect him to, even though there is no such thing as Santa Claus. Psychological verbs typically create intentional contexts.[20]

Because "S referred to Φ," "S expressed the idea of Φ," "S is thinking of Φ," and so on are intentional contexts, they do not express *relations* between S and Φ, and cannot be symbolized as two-place predicates in quantification theory. That is, they do not have the logical form "Rxy." The quantificational form "Rxy" is fully *extensional*, subject to both the substitutivity of identity and existential instantiation in both the "x" and the "y" position. We cannot be related to a thing in any way unless the thing exists, but we can think about and refer to things that do not really exist. Thinking of a winged horse is not like perceiving a winged horse or being a photograph of a winged horse. While the forms in question do not express relations between S and Φ, they do express properties of S, and can be symbolized as one-place predicates in quantification theory. They have the logical form "Px," and are fully extensional in the "S" position.

Intentional predicates lack other essential properties of relations. For example, if "Sam married *a woman*" is true, there must be a *particular* woman whom Sam married. He must have married Madonna, or Julia Roberts, or Margaret Thatcher, and so forth. But "Sam wants to marry a woman" may be true even though there is no particular woman whom

20 Brentano 1874; Chisholm 1955–6; 1958; Sellars 1958; Urmson 1968; Dennett 1969. Donnellan and many others have denied that speakers can refer to nonexistent objects. I argue against this position in *Nondescriptive Meaning and Reference*.

he wants to marry: he may not have found the right woman yet. Similarly, when Sam says "George hired a woman," he used the general term "a woman" to express the idea of a woman, but not to express the idea of any particular woman. Whereas relations are necessarily relations to particular objects, intentional attitudes may be *essentially indefinite*.

Causal theorists maintain that S can think about and refer to Φ only if Φ is causally connected to S's current mental state in a particular way.[21] Under the influence of this theory, many philosophers have begun to use the terms "about" and "of" relationally. This has led to some fallacies of equivocation that we need to guard against. Consider Putnam's (1981) famous "brain in the vat" argument against skepticism. Putnam has us imagine a brain isolated from any body since before birth. The brain is connected to an incredible life support system that provides not only sustenance but all of the neural inputs that would have been provided had it been in our body. While fantastically unlikely given the current state of technology, such a case contradicts no known psychological or physical law. Given that the vatted brain receives all the same inputs our own brain has had, we might think that it should think the same thoughts that we do. But the causal theorist denies this, on the grounds that the vatted brain's neural activity is not connected in the right way to the objects in its environment. Whereas we obviously can think about the Earth, the sky, and other things in our environment, Putnam believes, the vatted brain could not. This leads to Putnam's argument against the skeptical possibility that we actually are brains in vats:

(7) We can think about the Earth.
(8) If we were brains in vats, we could not think about the Earth.
(9) Therefore, we cannot be brains in vats.

In the intentional sense that I believe is conventional in standard English, (8) is false. Thinking about Φ does not even entail the existence of Φ, and so a fortiori does not entail being causally or otherwise related to Φ. But suppose we interpret the word "about" relationally, as the causal theorist does. On that interpretation, premise (8) may well be true. But then premise (7) would not state something we can know to be true solely on the basis of introspection. Indeed, premise (7) could not be accepted without begging the question against the skeptic. On the relational interpretation of "about," we cannot know the truth of (7) without perceptual

21 Cf. Harman 1973: §4.4; Devitt 1981: 79, 133; Lycan 1985: 87; Maloney 1989: Chapter 6; Fodor 1990b: 326; Sterelny 1990: Chapter 6: Fitch 1990; McGinn 1997: 80.

evidence that our thoughts are appropriately related to our hands, which would require our knowing that we are not brains in a vat. I will argue against the causal theory elsewhere.[22] What I wish to observe here is that I will always be interpreting "thinking about" and related expressions intentionally rather than relationally. As we shall interpret it, (7) expresses something we can know to be true solely on the basis of introspection, and (8) appears to be false, given what we know about the relation between the mind and the brain.

Whereas *speaker* reference is opaque and intentional, *word* reference is transparent and relational (§7.7). Since Mt. Everest is the tallest mountain, "the tallest mountain" refers to Mt. Everest. Mt. Everest is the referent or extension of that phrase. To say that "the golden mountain" does not refer to anything is just to say that there is no such thing as the golden mountain. This phrase has no referent. When Susie uses "Santa Claus," she is referring to someone even though she is mistaken in believing that the term she is using has a referent. A referring term is therefore one that can appropriately be represented by an individual constant serving as an argument of a predicate in classical quantification theory. Since "Fa $\therefore \exists$Fx" is a valid inference in quantification theory, the constant "a" must denote an existent object. Similarly, since "$a = b \therefore$ Fa $\equiv$ Fb" is valid in quantification theory, any context in which an individual constant appears must be extensional. The relationality and transparency of word reference distinguishes it in the same way from word meaning. "Pegasus" means "Pegasus," and therefore has a meaning even though it has no referent. "The morning star" and "the evening star" have the same referent even though they have different meanings.

§6.4 QUANTIFYING IN

An important case in which we will uniformly give a transparent interpretation to terms in normally opaque contexts arises when we use individual variables to state generalizations linking opaque contexts to transparent contexts. Consider first Theorem 3.5: *S meant that p by e iff S produced e as a direct and undisguised indication that he occurrently believes that p*. The letter "p" is a placeholder for sentences, and is thus a *substitutional variable*. Sentences are not names or singular terms; they express propositions and have meanings but do not refer to either. Substitutional variables in formal principles should be understood as bound by implicit universal

22 In *Nondescriptive Meaning and Reference*.

substitutional quantifiers. Theorem 3.5 asserts that all substitution instances of "S meant that p by e iff S produced e as a direct and undisguised indication that he occurrently believes p" are true. A substitution instance is any sentence of that form obtained by substituting a meaningful sentence for "p" or by assigning any sentence meaning to a constant replacing "p." Hence Theorem 3.5 entails sentences like the following for any S and e:

(10) S meant that Venus is the morning star by e iff S produced e as a direct and undisguised indication that he occurrently believes that Venus is the morning star.
(11) S meant that $2+2=4$ by e iff S produced e as a direct and undisguised indication that he occurrently believes that $2+2=4$.
(12) S meant that birds fly by e iff S produced e as a direct and undisguised indication that he occurrently believes that birds fly.

The symbols "S" and "e" are standard *individual variables*, which occupy places in formulas where singular terms can appear. Whereas substitutional variables do not range over a domain of objects, individual variables have the set of all objects in the domain of quantification as their domain. Individual variables should be understood as bound by implicit universal objectual quantifiers. So Theorem 3.5 should be understood as asserting the universal generalization of each of the above sentences (e.g., "$\forall S \forall e$[S meant that birds fly by e iff . . .]") and thus each of its universal instantiations.

Theorem 3.5 does not entail a false statement like the following, however.

(13) S meant that $2+2=4$ by e iff S produced e as a direct and undisguised indication that he occurrently believes that $(5-3)+(38 \div 19)=\sqrt{16}$.

Statement (11) does not entail statement (13) even though $2=(5-3)$, $2=(38 \div 19)$, and $4=\sqrt{16}$, because we intend "means that p" and "believes that p" to be opaque contexts. And (13) does not follow directly from Theorem 3.5, because no *uniform* substitution of sentences for sentence variables will produce it.

Now consider Definition 5.2: *S communicates with A iff S intentionally communicates something to A*. "S" and "A" are individual variables, bound by implicit universal objectual quantifiers, and subject to the standard rules of quantification theory, such as existential instantiation and substitutivity of identity. This is to be true despite the fact that "A" occurs in a context that is normally opaque. The "Φ" position is invariably transparent in

sentences of the form "S communicates with Φ" but is ordinarily opaque in "S intentionally communicates something to Φ." If General Rumstadt communicated with Fritz, and Fritz is the mole, then General Rumstadt communicated with the mole, whether he knew that Fritz was the mole or not. But if Rumstadt was not aware that Fritz was the mole, then he may have intentionally communicated something to Fritz without having *intentionally* communicated something to the mole. "Rumstadt intentionally communicated something to the mole" will be true on its transparent interpretation, but not on the opaque interpretation that we normally intend. However, as this example illustrates, "Rumstadt communicated with the mole" is true, because "Rumstadt intentionally communicated something to the mole" is true on its transparent interpretation. "S communicates with A" requires that S think of A in some way, but not in any particular way. Hence Definition 5.2 holds when both contexts in which "A" occurs are interpreted transparently.

Whenever we use individual variables in opaque contexts, they are to be interpreted transparently. We will give a "latitudinarian" interpretation to such transparent contexts. That is, we will adopt the following convention in our metalanguage.

6.3 **Convention:** *A normally opaque context "Σ[**x**]" containing an individual variable "**x**" should be understood as true iff "Σ[Φ]" is true for some "Φ" such that "Φ = x" is true.*

That is, $\mathbf{\Sigma}[\mathbf{x}] =$ df $\exists\Phi$ ($\mathbf{\Sigma}[\Phi]$ & $\Phi = \mathbf{x}$). Thus Definition 5.2 rules that "Rumstadt communicated with the mole" is true because "Rumstadt intentionally communicated something to Fritz" and "Fritz is the mole" are both true.

Since "$\Phi = \mathbf{x}$" is an identity statement, it entails "$\exists\mathbf{x}[\Phi = \mathbf{x}]$." So Convention 6.3 entails that "$\mathbf{\Sigma}[\Phi]$" must be true for some "Φ" such that "Φ exists" is true. Hence we will be interpreting quantificational variables in normally opaque contexts as fully extensional (cf. §6.3).

Observe finally that Theorem 3.5 mixes individual variables with substitutional variables. Whereas "p" is a substitutional variable, "S" and "e" are individual variables. So even though "S produced Φ as Ψ" is normally opaque in both the "Φ" and "Ψ" positions, the use of "e" signals a transparent interpretation. Hence Theorem 3.5 equates "S means that p by e" with "S produces Φ as a direct and undisguised indication that he occurrently believes that p, for some $\Phi =$ e."

We will generally use Roman letters as individual variables, and Greek letters as substitutional variables. Thus the occurrences of "Φ"

in Definition 6.1 are opaque, and the occurrences of "S" are transparent. Exceptions to the Greco-Roman rule include "p" and "q," which are so standardly used as placeholders for sentences rather than as individual variables that the chances of misunderstanding are minimal. When these are used to form that-clauses of propositional attitude ascriptions, they should be understood as opaque contexts.

Part Two

Languages and Semantic Acts

7

Languages

We now turn our attention from speaker meaning to word meaning. One of the elementary facts about word meaning is that words have different meanings in different languages. The written word "rot" means red in German and decay in English. "Hood" means the engine cover in American English and the top of the car in British English. So we need to say something first about what languages are. As David Lewis (1975) has observed, *languages* must be distinguished from *language*. Languages are things like French, German, and English. French is *a* language, but is not language itself. Language is a human activity, in which languages are used. It includes speech, writing, and other types of symbol use. While linguists and philosophers sometimes debate as to which is primary, it should be clear that languages and language are complementary subjects.[1]

We will first examine what a language is, and then analyze what it is to use one. This will enable us to define what it is for a word to mean something on a given occasion ("applied" word meaning). We will be using "languages" in a very general sense, to denote symbol systems of all kinds, including codes, signal systems, sign languages, and artificial languages. The distinctive features of living, natural languages will be discussed in Chapter 11, culminating in a definition that implicitly defines what it is for a word or other linguistic unit to express or mean something in a living language. As far as possible, we will ignore those features of language that are irrelevant to meaning.

1 See Bentham 1843: 298; Fodor and Katz 1964: 1–19; Searle 1969: 18–19; R. Harris 1980: 22–32; Leech 1983: esp. Chapter 1; Avramides 1989: 6. Contrast Harrison 1980: 165–7 and Mey 1993: §1.3.

Languages on our view are systems of "modes of expression." While languages may be defined in part by rules for using words to express other mental states, we will be focusing on ways of expressing ideas. We will observe that natural languages such as English have secondary rules by which words can be used "ideo-reflexively" to refer to ideas by expressing them, giving us expressions like "the idea of water" and "the thought that water is wet." We need to understand ideo-reflexivity in order to understand the central claim that a word means "water" iff it expresses the idea of water.

§7.1 LANGUAGE MODELS

Languages are abstract objects of a certain sort. They are sometimes defined as sets of sentences or expressions.[2] But it is possible for there to be a code whose sentences are those of English, but in which all of the sentences have a meaning different from the meaning that they have in English. Indeed, a speaker of English could use such a code. If languages were identified with sets of sentences, such a speaker would be using English even when he was speaking in his code.

A language must therefore be an abstract object that associates meanings with words or other symbols. Lewis has suggested that a language is a *function*, whose domain is the set of all (grammatical) expressions in the language, and whose range is the set of their meanings. But this suggestion runs afoul of the fact that few, if any, expressions in natural languages have unique meanings. To handle ambiguity, therefore, Lewis identified languages with functions from expressions to sets of meanings. An alternative is to identify a language with a *relation*. English, on this suggestion, is the set of all ordered pairs whose first elements are the grammatical expressions of English, and whose second elements represent their meanings. A language in which some expressions are ambiguous is a one-many relation. A language with some synonyms is a many-one relation. All natural languages are both one-many and many-one relations.[3]

According to Lewis's specific proposal, languages are functions from *sentences* to *sets of possible worlds*. The possible worlds are those in which the sentence is true in the language, and they collectively represent the meaning of the sentence. On this approach, however, all of the sentences of the language that express a necessary truth come out being synonymous.

2 See, e.g., Bentham 1843: 303; Mates 1950: 111; Chomsky 1965: 15–18; and Loar 1981: 222–3.

3 Cf. Linde and Labov 1975: 925; Smith & Wilson 1979: 172; and Cresswell 1985: 9–10.

Yet "2 + 2 = 4" means something radically different from "Every triangle is either equilateral, isosceles, or scalene," at least in English. Moreover, the words, phrases, and other subsentential units of the language are assigned no meanings. These problems can be avoided quite naturally by letting the first element of a language relation be any grammatical expression of the language, and letting the meaning elements be *ideas* rather than possible worlds. The idea that 2 + 2 = 4 is radically different from the idea that every triangle is either equilateral, isosceles, or scalene. So a relation that assigns these ideas to different sentences will assign them different meanings. And a language relation can assign a meaning to a subsentential unit like "red" by assigning it the idea of red. We can then add that a sentence is true in a given language at a given possible world provided that the idea assigned by the language to that sentence is true in that world. So the meaning assignments determine truth conditions.[4] In a few cases, such as interjections, the meaning elements are mental states other than ideas.

Taking a language to be a relation associating ideas (or other mental states) with expressions fits nicely, of course, with the analysis of speaker meaning given in Part I. What a speaker means is generally determined by what ideas the speaker expresses. If we can show how the ideas that speakers of the language use the words to express determine the ideas assigned to expressions by the language, we will have shown how speaker meaning determines word meaning. We will do that in Chapters 8–11.

The nature of ideas will be clarified in Part III, and the choice of ideas as meaning elements will be defended against common objections in Part IV. Briefly, *ideas* (or, equivalently, *concepts*) are event-types of a certain sort, specifically, thoughts or parts of thoughts. Hence ideas are mental representations. The change in a man who is thinking "All cats are mammals" one minute and "All dogs are mammals" the next is a change in the ideas that are occurring to him. The elementary ideas that changed are both occurring to a woman who is thinking the thought that all cats are dogs. The same goes for the thought that all dogs are cats, which is a different structured event-type in which the same component ideas stand in different relations to each other. Some thoughts are propositions, and are the objects of belief, desire, intention, and the other propositional attitudes. To express an idea is to perform an observable action with a certain intention, roughly, the intention of providing an indication that the idea is occurrent (Chapter 3).

4 Cf. Loar 1981: 153, 220, 224–5; Fodor 1981: 196; Bach 1987a: 62–4; and §23.3, this volume.

Ideas in the sense indicated are to be sharply distinguished from two other sorts of mental representations, namely *images* and *conceptions*. Images are complexes of sensations (§19.2, §19.3). Conceptions are belief systems (§19.4). The images and conceptions associated with an expression are important determinants of its usage. Nevertheless, meaning is not determined by either, and does not consist in their expression. First of all, the association between words and images is not properly described as "expression." We do not *express* images, and the word "cat" does not express the image of a cat. What expresses our conception of gravity is not the word "gravity" but our complete theory of gravitation. Secondly, meanings do not stand in a one-to-one correspondence with either images or conceptions. (1) *Conceptions or images may be identical when meanings differ.* Our conception of a three-sided polygon is identical to our conception of a three-angled polygon, and any image we form of the former is an image of the latter. But the meaning of "three-sided polygon" differs from that of "three-angled polygon." Similarly, any image that would fit "the red ball" would equally well fit "The ball is red," despite the marked difference in meaning between phrases and sentences. (2) *Meanings may be identical when conceptions or images differ.* There have been changes in our conception of Earth over time, which were not paralleled by changes in the meaning of "Earth." And different English speakers invariably form different images, and have somewhat different conceptions, of Earth. Since the thought that three-sided polygons are three-angled polygons differs from the thought that three-sided polygons are three-sided polygons in containing different subject and predicate concepts, the idea of a three-sided polygon must be different from the idea of a three-angled polygon. The idea of Earth occurred to both Ptolemy and Copernicus despite their different conceptions of Earth, and despite different images of Earth they may have formed. Last but not least: (3) *Many meaningful expressions are not associated with images or conceptions, such as "is," "if," "and," and "the."* While sentences are associated with the same images as related phrases, it is hard to figure out what conception (belief system) to associate with a single sentence, such as "Earth revolves around Mars." The idea that the meaning of a sentence is determined compositionally by the meanings of its words becomes positively mystifying if we interpret meanings in terms of conceptions (see §19.4). So expression theories of meaning focusing on the expression of thoughts and thought parts do not have the obvious defects of image and conception theories.

It is unnecessary and inaccurate to refer to ideas (or sets of possible worlds) as meanings (see §21.1). The word "red" means *red* in English.

It does not mean *the idea of red*. Ideas occur to people. Meanings do not. Ideas can be true or false, but not meanings. Unlike "expresses," the verb "means" does not express a relation in English (see §6.2 and §6.3). In viewing languages as relations between expressions and ideas, we are construing languages as expression relations, not meaning relations.

The claim that languages *are* sets of ordered pairs is also problematic. It implies that languages have members and subsets, which sounds like a category mistake. Moreover, to defend the claim, we would have to provide a reason for thinking that languages consist of expression-idea pairings rather than of idea-expression pairings. But that choice was completely arbitrary. While ⟨e,i⟩ and ⟨i,e⟩ are different ordered pairs, it makes no difference whether we construe a language as the set of all pairs ⟨e,i⟩ such that e expresses i in the language, or as the set of all pairs ⟨i,e⟩ meeting the same constraint. Finally, languages change over time, and could have been different. These things cannot be said of sets. What we can justifiably say is that expression-idea pairings are language *models*. Language models *represent* the relevant properties of languages in a perspicuous way that facilitates formal treatment. The fact that there are other equally good models is insignificant.

§7.2 SYSTEMS OF MODES OF EXPRESSION

A language is a system of a certain sort, a combination of elements. We have ruled out expression-idea pairings as the elements of the system. What is the alternative? We have defined expression in terms of intention, and defined speaker meaning in terms of expression. It is a platitude that there are many different *ways* of expressing ideas, which may vary from language to language. Expressing the idea of red by using "red" is one way. Using "rot" to express the same idea is another. Ways of expressing ideas are themselves semantic acts-types, specific determinates of the general determinable expression. Languages may be characterized as *systematic ways of expressing ideas and other mental states*. More generally, languages are systematic ways of using expressions. To learn, teach, use, or create a language is to learn, teach, use, or create a systematic way of expressing ideas. Languages arise and change as systems of expression arise and change.

Let us use *mode of expression* as a technical term to denote *the use of a particular expression to express a particular idea or other mental state.* A mode of expression is direct or indirect depending on whether the idea is expressed directly or indirectly. Modes of expression are not only ways of expressing

ideas but also ways of using expressions. Expressing the idea of red by using "red" is as much a way of using "red" as it is a way of expressing the idea of red. For every mode of expression there is a unique ordered pair whose first element is an expression and whose second element is an idea (or other mental element). Hence ⟨"red," the idea of red⟩ represents a mode of expression.[5] Whereas ⟨e,i⟩ and ⟨i,e⟩ are different ordered pairs, using e to express i and expressing i by using e are one and the same mode of expression. Languages may now be defined as systems of modes of expression.

7.1 **Definition:** *A language L is any system of modes of expression.*

Since a mode of expression can be represented by an ordered pair ⟨e,i⟩, a language can be represented by a language model.

Not all ways of expressing ideas that speakers of the language may use are parts of the language. A speaker of English may express the thought that she cannot go to the movies by uttering "I have to study." But this way of expressing ideas is not essential to English, and English would not be any different if there were no such way to express that thought. The fact that this mode of expression is indirect is not decisive (§2.3). For some modes of indirect expression are constitutive of a language. For example, it is constitutive of English that "It will rain or snow" can be used to directly express the disjunctive thought "It will rain or snow," and that involves indirectly expressing the component thoughts "It will rain" and "It will snow." The difference between the two cases is that the indirect mode of expression is entailed by the direct mode in the disjunction case but not in the relevance implicature case. That is, one can use "I have to study" to express what it directly expresses in English without using it to express the thought that one cannot go to the movies. But one cannot use "It will rain or snow" to express what it directly expresses in English without using it to express indirectly the thoughts "It will rain" and "It will snow." Since the indirect modes of expression constitutive of a language are those entailed by the direct modes of expression constitutive of that language, languages are defined by a set of direct modes of expression.

A system is a combination of elements, but not every combination is a system. So Definition 7.1 does not imply that every arbitrary assemblage of modes of expression counts as a language. It would certainly

5 My "mode of expression" thus resembles Saussure's (1916: 67) "sign" (the "combination" of a signifier and a signified) and the notion of "symbol as sign together with sense" that Carruthers (1989: Chapter 4) attributes to Wittgenstein.

be counterintuitive to think of the sum of the two modes of expression represented by ⟨"red," the idea of red⟩ and ⟨a certain motion of the hands, the idea "Safe!"⟩ as a language. To constitute a system, there must be some unifying, structural features relating the elements combined. In the case of natural languages, the fact that all expressions are sequences of a small set of speech sounds or letters is one of many unifying features. Another is the fact that the indirect modes of expression in a language are entailed by the direct modes. Since the systematic nature of languages is not essential to the fact that words have meaning in them, we will not analyze it further, and will place no theoretical weight on it. The definitions and theorems given here could all be generalized, in fact, by letting "L" range over "symbol complexes" (any combination of modes of expression, systematic or not) rather than over languages. Alternatively, Definition 7.1 could be narrowed by modifying "system" with a restrictive adjective without undermining anything we have to say.

Since natural languages are continually evolving and artificial languages are frequently invented, it is necessary to assume that some languages have not been *established* or *created*. A system of expression is established through *use*, and created by *stipulation*. Section 7.5 will define what it is to use a language, and Chapter 11 will develop the idea that natural languages become established through *conventional usage*. Dead or extinct languages were established at some time in the past but are no longer used today.

§7.3 WORD MEANING AND EXPRESSION

The concept of a mode of expression is defined in terms of *speaker* expression. A mode of expression is a way of expressing an idea or other mental state, that is, a way for a speaker to express something. We will now define word expression in terms of speaker expression by defining it in terms of mode of expression. Let "e" and "i" be general individual variables. The former will generally occur in formulas where terms referring to expressions would occur, and the latter in formulas where terms referring to ideas would occur, but we will not stipulate a restrictive range for these variables.

7.2 **Definition:** *e expresses i in L iff using e to directly express i is one mode of expression in L.*

While Definition 7.2 does define word expression in terms of speaker expression, it does not answer the fundamental question that began our inquiry. It sheds no light, for example, on what it is for the word "red" to

express the idea "red" in English. Definition 7.2 does tell us that in order to answer the fundamental question, we will need to identify what makes it true that the use of "red" to directly express the idea of red is one mode of expression in English. We will answer that question in Chapters 8–11.

In English, "express" is ambiguous when applied to words and other expressions. There is an inclusive sense in which "vixen" expresses the idea of a fox and the idea of a female, as well as the idea of a female fox. And there is a stronger, exclusive sense in which "vixen" expresses only the idea of a female fox. In the exclusive sense, it is "fox" rather than "vixen" that expresses the idea of a fox. Definition 7.2 defines the exclusive sense of word expression, and that is the one we will use throughout.

It is customary to refer to L, the language that we are studying, as the *object language*. The language that we are using to describe L is the *metalanguage*. The metalanguage in this work is obviously English – augmented by technical terms and modified by restrictive stipulations. The object language will usually be English, but not always. When there is no explicit reference to a language, the implicit reference is to the language that is the contextually indicated object of discussion.

7.3 **Definition:** *e expresses i iff e expresses i in the object language.*

In other words, "e expresses i" *simpliciter* makes an indexical reference to the language under discussion. Hence we say " 'Hood' expresses the idea of the engine cover," with no explicit reference to a language, when we wish to say that "hood" expresses the idea of the engine cover in the language we are talking about. The statement is accordingly true when the object language is American English, false when it is British English. This holds whether American English or British English is our metalanguage.

Ideas are not meanings, but they are *meaning determiners* or *meaning elements*. What a word means in a language can be defined in terms of what it expresses in that language. As in §2.4 and Table 7.1 (p. 142), let "μ" be a placeholder for the *oratio recta* expressions that can follow "means" and "idea." Let "μ_i" be a more specific placeholder standing for linguistic units expressing i at the place occupied by "μ_i" in a formula. Definition 7.4 uses these conventions to define word meaning in terms of expression.

7.4 **Definition:** *e means μ_i (in L) iff e expresses i (in L).*

Thus a word *means* "red" in English iff it *expresses the idea* "red" in English. "Rot" means "red" in German because it expresses the idea of red in German. "C'est rouge" means "That is red" in French because

"C'est rouge" expresses the idea (thought, proposition) "That is red" in French. In making these claims about English, German, and French, which are our object languages, we are using English, which is our metalanguage. We have been using "red" in this paragraph to express the idea of being the color red, rather than the idea of being a communist. "Red" is an expression of the metalanguage, which we are using to describe the meaning of "rot" and "C'est rouge" in different object languages. " 'Rot' means 'red' in L" says nothing about what "red" means in L, or about what any word means in the metalanguage (unless we happen to be using L).[6]

Definition 7.4 is intended to hold when the parenthetical material is absent as well as when it is present, so it follows that what an expression means *simpliciter* is what it means in the object language. Thus *"Rot" means "red"* must be counted as false when we are describing English, even though "rot" means "red" in German. When German is the object language, the same sentence is true. Of course, since the metalanguage usually is the object language – we are normally most interested in the properties of the language we are using – it is also true that "e means μ" is normally true when e expresses the idea μ in the metalanguage. Exceptions arise when we are talking about a foreign language. Thus when our object language is German, *"rot" means "red"* is true even though "rot" does not mean "red" in our metalanguage. Note that we can use foreign words in our metalanguage. Thus when translating English for a Frenchman, I might say " 'Dog' means *chien*." Then I used "chien" to express the idea of a dog, and I said that "dog" expresses that idea. The implicit object language is English, and my metalanguage is a mixture of English and French.

We can go on to say that two expressions mean the same thing in a language (they are *synonymous*) if the language assigns them the same idea, and that one word means two different things (it is *ambiguous*) if the language assigns it two different ideas. The word means something, has a meaning, and is *meaningful* if it expresses something. *The meaning of e is* μ if μ is what e means, that is, if the idea μ is expressed by e. We say that the meaning of "rot" is "red," for example, because "rot" means "red." The meaning of e is a compositional function of the meaning of its parts if the idea expressed by e is composed of the ideas expressed by its parts (Chapter 10 and §14.2). None of this requires holding that ideas are meanings, or that "means" is a relational predicate. When e means μ, the meaning that e has can be identified with the feature that e has of

6 Cf. Chisholm 1958; contrast Sellars 1958.

meaning μ (see §21.1). Given Definition 7.4, meaning μ_i is the property of expressing i. Hence the meaning that "rot" has is not the idea that it expresses, but the feature that it has of expressing that idea. Consequently, for every ordered pair $\langle e,i \rangle$, there is a unique ordered pair $\langle e,m \rangle$ in which m is the feature of meaning μ_i. If we call expression-idea pairings *deep* language models, and expression-meaning pairings *surface* language models, then we may say that there is a one-to-one correspondence between deep and surface language models.

We are focusing on the overwhelmingly typical case in which words have meaning because they express ideas. As we have observed before (§2.5), interjections are exceptions to this rule. "Aïe!" means "Ouch!" in French, but we cannot say that "Aïe!" expresses the idea "Ouch!" "Ouch!" does not express an idea, so "the idea 'Ouch!' " does not have a referent. These facts do not falsify Definition 7.4, because we have not stipulated that "i" must denote an idea. Thus Definition 7.4 will count " 'Aïe!' means 'Ouch!' in French" as true as long as "Aïe!" expresses the mental state in French that I am using "Ouch!" to express in English. While "Ouch!" does not express an idea, it is used to express at least one mental state, namely pain.[7]

In defining languages as systems of expression, I am not denying that words are used to perform actions other than expression. It is undeniable, for example, that "Hello!" is used to greet people and "Go home!" is used to order people home. "It is going to rain" is used to assert, predict, and warn that it is going to rain. What I am assuming is that in all such cases, we use words to perform actions by using them to directly express ideas or other mental states. It would be easy enough to define languages as modes of use, the bulk of which are modes of expression. But while theoretical options are available, I believe expression is the fundamental form of language use. Reasons for believing this even in the case of words like "Hello!" will be presented in my forthcoming *Nondescriptive Meaning and Reference.*

§7.4 LINGUISTIC VARIATION

We have thus far ignored several complexities. First, natural languages exhibit great *variation* at any given time, and have distinct regional and social varieties called *dialects*. Thus American English and British English are dialects of English. In addition to differences in spelling and

7 Interjections will be discussed more fully in *Nondescriptive Meaning and Reference.*

pronunciation, hundreds of words have different meanings in the two dialects, such as "bonnet" and "hood." Black American English and Midwestern American English differ in grammar as well, with sentences like "He thinking about it" and "The teachers don't be knowing the problems" being acceptable in the former but not the latter (Fasold 1990: 208). Chinese has dialects such as Cantonese and Mandarin that are mutually unintelligible except in writing.

It should be emphasized that dialects are themselves languages: "is a dialect of" expresses a relation between two languages. Both American English and English are systems assigning meanings to expressions. Furthermore, I use the term "dialect" descriptively, to denote any variant of a given language, with no connotation that the variant is in any way substandard (see §9.6). Given Definition 7.1, the relation between a language and its dialects is the relation between a system and its subsystems, that is, between a whole and its parts. The use of "hood" to mean "the top of the car" is part of British English but not American English, while the use of "hood" to mean "engine cover" is part of American English but not British English. The set-theoretical model of a language is the *union* of the models of all its dialects, which are its model's *subsets*. The ordered pair ⟨"hood," the idea of the top of the car⟩ is in the model representing British English but not American English, while ⟨"hood," the idea of the engine cover⟩ is in the model representing American English but not British English. Both are in the model representing English.

British English and American English are not dialects of each other. Instead, they are dialects of a common and more inclusive language. Two languages *overlap* to the extent that they share ways of expressing ideas. American English and British English overlap almost completely, while Cantonese and Mandarin Chinese overlap only in writing. It is important to note that there need be no *linguistic* reason why one language is considered a dialect of a second rather than of a third. For example, Low German is a dialect of German despite being more similar linguistically to certain dialects of Dutch than to some other dialects of German.[8] It is also hard to see much difference between dialects and *styles* or *registers*, except that one and the same individual switches from style to style as the occasion demands, no matter what dialect he or she speaks (Labov 1970:

8 Lyons 1971: 35; Anttila 1989: 182–3, 289–9; Vogelin et al. 1988: 659ff.; and J. Simpson (1994: 1895). Simpson hastily concluded that "a language" is not definable in linguistic theory because "no firm boundary can necessarily be drawn between one language and another and one language may contain vast differences of pronunciation, grammar, and vocabulary" (1994: 1896).

Chapter 3). There are numerous differences in pronunciation, vocabulary, and grammar depending on whether the style is formal or informal. "Going to" is pronounced "gonna," contractions and slang are permitted, and "tick off" is used to mean "make angry" only in informal English. Styles are situational rather than regional or class variations.

The second complexity is that languages are capable of *change*, which is variation over time. They are temporally extended systems. Natural languages change constantly (if slowly) with regard to vocabulary, grammar, and meaning. They have an origin, and eventually evolve into other languages or become extinct. At one time "I've" meant "female horse" in French, later "cavale" had this meaning, and today "jument" does. "Jument" at one time meant "pack horse" before coming to mean "female horse" (Anttila 1989: 134). Between the eleventh and fourteenth centuries, the distinction between the nominative and the oblique cases disappeared in French, and the final "s" came to be the mark of the plural (Lyons 1971: 47). Words are constantly added to one language by "borrowing" them from neighboring languages, or from classical sources such as Latin and Greek. Scientific and technological terms are being added at a rapid rate. Each generation invents its own slang, which normally obsolesces rapidly. Linguists have estimated that basic vocabulary items disappear from a language at an average rate of 81 percent per thousand years (Woodward 1978: 337; Anttila 1989: 396). The growth in the total word stock of English has been exceptional, moving from 100,000 words in the fifteenth century to over 450,000 today (Landau 1984: 37).

As for evolution, the etymology of a word traces its history. According to *Webster's Encyclopedic Unabridged Dictionary*, the modern English word "disease" comes from the Old English "disese," which derived from the Old French "disaise" ("dis" + "ease"). More globally, Old English changed into Middle English. French, Italian, Spanish, Portuguese, and Rumanian all descended from Latin. Language *families*, such as the Romance languages just mentioned, are groups of languages with a common origin. The dialects of a language also have a common origin. Isolated languages that share an unusual feature are assumed to have had a common origin, in which the feature arose just once (Anttila 1989: 302–3). Different dialects may evolve at different rates in different respects (Fries 1940: 52, 98; Fasold 1990: Chapter 8).

While languages change, and wholes change their parts, sets of ordered pairs do not. Sets of expression-idea pairings therefore model *language states* rather than languages. Languages as temporally extended systems may be

modeled by *functions from times (or time intervals) to language state models*. Thus if $\mathcal{L}(t)$ represents a language L, then $\mathcal{L}(t_1)$ is a set of expression-idea pairs, $\mathcal{L}(t_2)$ is a set of expression-idea pairs, and so on for all times t. L changes between t_1 and t_2 iff $\mathcal{L}(t_1) \neq \mathcal{L}(t_2)$. Thus French at an early date contained the mode of expression ⟨"I've," the idea of a female horse⟩, at a later date ⟨"cavale," the idea of a female horse⟩, and today contains ⟨"jument," the idea of a female horse⟩. The description of languages at a particular time, called "synchronic" description, is the task of descriptive linguistics. The description of languages as they change over time, called "diachronic" description, is the subject of historical linguistics. Genetic or comparative linguistics studies the evolution of similar languages from common ancestors.[9]

Finally, there is modal variation. Natural languages *might have been different* from the way they actually are. For they might have evolved in a different way. Thus language models must be functions from times *and possible worlds* to sets of expression-idea pairings. Definition 7.2 should be understood as having an implicit quantification over times and possible worlds. That is, Definition 7.2 is short for *e expresses i in L(t,w) iff using e to directly express i is one mode of expression in L(t,w)*. Other references to L in this chapter should be understood similarly. We will normally be concerned with the present time and the actual world.

§7.5 LANGUAGE RULES

Each mode of expression ⟨e,i⟩ in a language L represents a *rule* specifying that it is permissible in L to use e to express i and mean μ_i. Indeed, we could use sets of such rules as our language state models, with only a loss of some convenience.[10] Such rules may be called the *material* or *constitutive* rules of a language. They do not tell us anything directly about the specific formal features of the language, and yet they constitute the facts on which all *structural* rules must be based, directly or indirectly. The

9 Cf. Lyons 1977: 243: "If we apply strictly the distinction of the diachronic and the synchronic, we will say that the notion of one language (e.g. English) existing over the centuries (from the time of Shakespeare to the present day, shall we say) is fallacious. What we have underlying the language-behaviour of people living at different periods are distinct language-systems; each of these systems can be studied, synchronically, independently of the other; and diachronic linguistics can investigate how an earlier system was transformed into a later system." As I hope I have made clear, Lyons's first claim does not follow from the truth of his second claim.

10 Alternatively, a language can be modeled by a set of rules (comprising a "grammar") that generate all and only the material rules of the language, as suggested by Katz (1966: 100–5, 112).

structural rules may be highly complex universal statements, or more or less rough generalizations with known exceptions.

Among the *phonological* rules of a language are statements describing each spoken expression of the language (if there are any) as a sequence of elementary speech sounds. Other phonological rules describe each speech sound in terms of a small set of phonetic or articulatory features, and assign it to a phoneme. The *orthographic* rules systematically correlate any written expressions with spoken expressions. As a first approximation, there are three basic types of writing schemes, depending on what units of the spoken language the basic units of the written language correspond to. The elementary written expressions of the language correspond (more or less roughly) to the sounds of the spoken language in alphabetic systems such as English, to syllables of the spoken language in syllabic systems such as Arabic, and to whole words of the spoken language in ideographic systems such as Chinese.[11] Whatever the type of system, the rules are such that if e_1 is written e_2, then e_2 is pronounced e_1, and conversely. Moreover, e_1 and e_2 express the same idea in the language: mode of expression $\langle e_1,i\rangle$ is in L iff $\langle e_2,i\rangle$ is. Thus the material rules for the written expressions of L can be derived from the material rules for the spoken expressions together with the orthographic rules.

A modern language such as English may be viewed as the combination of a spoken language and a written language. The largely derivative character of writing has led some to deny its status as language altogether.

> The definition of language often specifically excludes writing, because it is a secondary representation of the primary speech. Furthermore, writing is a very recent innovation and is not a necessary accompaniment to language. (Anttila 1989: 31)[12]

> Writing differs from language . . . in a very fundamental way. Language is a natural product of the human mind – the properties of people that make it possible for everyone to learn any language, provided they start at a young enough age – while writing is a deliberate product of human intellect: no infant illiterate absorbs its script along with its language; writing must be studied. Language continually develops and changes without the conscious interference of its speakers, but writing can be petrified or reformed or adapted or adopted at will. (Daniels and Bright 1996: 2)

11 For a more complete and accurate typology, see Daniels and Bright 1996: 4, 6, 9.

12 See also Saussure 1916: 23–5; Bloomfield 1933: 21; Hockett 1958: 4; 1966: 14–15; Algeo 1973: 17. Contrast R. Harris 1980: 6–18; S. Landau 1984: 207; and Robins et al. 1997: 557.

These facts do show that written expressions provide a poor sample for studies aimed at discovering linguistic universals. But they do not begin to imply that writing is not part of the phenomenon known as language. Written language is language. In writing this book, I am using language. I am using a particular language, English. There are words on this page, and they are words of English. The sentences I have written are meaningful, and most of them, I hope, are true.

Nor is speech a necessary accompaniment of language: languages can be invented and used that contain nothing but written symbols. American Sign Language is one of many native languages that have developed among the deaf in which ideas are expressed by signing rather than by speaking or writing.[13] Deaf children in deaf families learn American Sign Language in the same automatic way that hearing children learn spoken languages. Moreover, the fact that the spoken component of a natural language may be historically, developmentally, and explanatorily primary does not entail that the written expressions are not even part of the language. Since the alphabetic sequence "rot" means something in English, and something different in German, it must be reckoned an expression of both languages. What is crucial for our purposes is that *written words have meaning in exactly the same sense as spoken words*. So do the signs in American Sign Language. The alphabetic sequence "Snow is white" is just as meaningful and just as true as the series of speech sounds that it represents. Hence no restrictions should be placed on the physical medium of the meaningful expressions in a language. The expression variable "e" ranges over anything that can be used for the purposes of expression – anything by which someone might mean something.

The paradigm structural rules of a language are its *grammatical* rules, which describe directly or indirectly how simpler meaningful expressions in the language can be combined to form more complex expressions of the language. Thus "the," "cat," and "died" are expressions of English, as is "the cat" and "The cat died." But "cat the" and "Died cat the" are ungrammatical and have no meaning in English. In linguistic theory, a *grammar* for a language is a system of rules that generates all and only the grammatical expressions of the language, along with a structural description of each. The structural description describes the units of which the expression is composed, the manner of their combination, the formal relations of the expression to other expressions, and so on. In a transformational grammar, the structural description contains a *surface structure*

13 See Deuchar 1984: 24; Kyle & Woll 1985; Wilbur 1987; Schein & Stewart 1996: ix.

and a *deep structure*.[14] The surface structure describes how the expression is composed of smaller expressions, assigning each to a grammatical category. The deep structure depends on the meaning of the expression, and represents grammatical information not contained in the surface structure. For example, "John is easy to please" and "John is eager to please" have the same surface structure, partially represented by "Noun is adjective to verb." But the former is related to "Others please John" in a way that the latter is related instead to "John pleases others." A major issue in linguistic theory is whether there are purely syntactical deep structures separate from the ideas or semantic structures expressed by expressions, and, if so, how the syntactic and semantic structures are related.

On some approaches,[15] a grammar also has a semantic component, which assigns a meaning to each expression on the basis of its structural description. Such a grammar is complete only if it generates all the expression-idea pairings constituting the language. On other approaches,[16] not at all incompatible with the preceding ones, the ultimate deep structures are thoughts and ideas. The grammar encodes the thought, and the surface structure is derived from that grammatical representation. In other words, a semantic theory may be *encoding* ("generative") or *decoding* ("interpretive"). In the decoding type, a sentence and its structural description provide the input to a semantic theory, and the output represents the idea expressed by the sentence. On the encoding model, an idea and its structural description provide the input to the semantic theory, and the output is a sentence expressing that idea. Looked at in another way, each language determines a relation between expressions and ideas. Each such relation determines two functions. The *encoding function* associated with a language assigns to each idea the unique set of expressions that express that idea in the language. The *decoding function* assigns to each expression the unique set of ideas that it expresses in the language. Semantic theory might seek a finite, formal specification of either the encoding or the decoding function.

Of special importance here are *constructive semantic* rules, or *construction* rules, which state how the idea expressed by a complex unit is determined

14 Or sequence of deep structures. Cf. Chomsky 1964: 51–64; Postal 1964; Fodor & Katz 1964: 153–4.

15 Chomsky 1964: 51–2; 1975: 41; Katz and Fodor 1963: 503; Katz 1966: 151–85; 1972; 1977a; Fodor 1975: 151.

16 Harman 1972; 1973: Chapter 5; McCawley 1968; 1994; Quillian 1968: 244; G. Lakoff 1971a; Bennett 1976: 252–3; Schank and Abelson 1977: 7–8; and Chierchia and McConnell-Ginet 1990: 8, 144–5.

by the ideas expressed by its components and their mode of combination. For example, reflection on the meaning of phrases like "man hater," "cat lover," "piano player," "aircraft carrier," and the like reveals a general pattern. If N stands for a noun, and V for a transitive verb, a phrase of the form "N V-er" expresses the idea of something that Vs Ns rather than the idea of a V-er that is an N. Such rules implicitly define the meaning of the *constructions* found in the language, such as "N V-er." If the material rules of a language that assign meanings to grammatically atomic expressions are called *lexical* rules, then the lexical rules combined with the constructive semantic rules entail all of the material rules of a language that assign meanings to grammatically compound expressions.[17] The rules assigning meanings to idioms may for this purpose be reckoned either as lexical rules, or as construction rules with only one instance.

The *referential* or *extensional* properties of the expressions in a language, along with their *logical* properties, are derived from those of the ideas they express (§23.3). For example, "There are more than a billion people in the world" is true as interpreted in American English because it expresses the thought in American English that there are more than 10^9 people, and there are in fact more than 10^9 people. The same sentence is false as interpreted in British English because it expresses the thought in British English that there are more than 10^{12} people, and there are not in fact more than 10^{12} people.

§7.6 RULES FOR IDEO-REFLEXIVE REFERENCE

Referring to an idea and *expressing* it are different semantic acts. To refer to an idea is to express the idea *of that idea* (Chapter 6). We express the idea of a horse when we are talking about horses, as when we say "Secretariat was a magnificent horse." We refer to the idea of a horse when we are talking about the idea, as when we say "The idea of a horse is occurring to me now." As this example illustrates, languages like English have productive rules by which we can refer to an idea by using in certain contexts a word that expresses the idea.[18] *The idea of a horse* refers to the idea that *horse* expresses. Similarly, *the idea "red"* refers to the idea that the word "*red*" expresses. The expression "c(red)," which we will occasionally use

17 Katz & Fodor 1963: 493; Katz 1966: Chapter 4; Lewis 1969: 168–71.

18 Cf. Kneale and Kneale 1962: 585; Woodfield 1982: 281; Crimmins 1992: 163–6; and Horwich 1998a: 26. Compare and contrast Sellars 1963: 7; 1979: Chapter 4; Davidson 1968; Peacocke 1975: 4; Lycan 1981, 1985; Böer & Lycan 1986: Chapters 3–4; Forbes 1993; Heal 1997: 638–9. See also *Nondescriptive Meaning and Reference.*

Table 7.1. *The placeholder "μ"*

means μ	*idea μ*	*μ*
means "plane"	idea "plane"	"plane"
means *plane*	idea *plane*	*plane*
means plane	idea plane	plane

as an abbreviation for "the concept of red," is understood in a similar fashion. The convention in cognitive psychology is simply to italicize the words, or put them in small caps, in order to refer to the concepts they express. Thus we can say that SALT activated PEPPER, meaning that the one idea activated the other (as in §18.2). I call this the *ideo-reflexive* use of words. The rules of ideo-reflexive usage enable the word-idea pairings determined by a language to serve two different purposes, and thereby increase the efficiency of the language as a means of expression.

We have been using "μ" as a placeholder for words used ideo-reflexively. Two of the principle contexts are illustrated in Table 7.1 (which recalls Table 2.1).

Like "red" and "plane," most words are ambiguous, and therefore express more than one idea in English. On any occasion on which "the idea 'red' " is used, however, the word "red" is used with a particular meaning. "The idea 'red' " is used to refer to the idea that "red" expresses on that occasion. Thus when we are talking about colors, "the idea 'red' " designates the idea of being the color red. When we are talking about political orientations, the same phrase refers to the idea of being a communist. The notion of what an expression means or expresses on a given occasion is defined in §7.9.

In general: *on any occasion of use, "the idea μ" refers to the idea expressed by "μ" on that occasion, if there is one.* I will show in *Nondescriptive Meaning and Reference* that this rule holds because expressions of the form "the idea μ" are indexical descriptions used deictically. A substitution instance of "the idea μ" has the referent that it has because the speaker is aware of the idea expressed by "μ"on that occasion. "Ideo-reflexive" indexicality thus resembles the more familiar "token-reflexive" form in that the speaker refers to something present that is related to the word used, but differs in what is presented and the way it is related to the word used.

To capture the rule governing "the idea μ" formally, we need to index "μ" with variables and constants standing for ideas or other expressibles. That is, we need to use placeholders of the form "μ_i," which stand for linguistic units expressing a particular idea i at the place occupied by

"μ_i" in a formula. If *a* is the idea of an airplane, then "means 'plane'" is a substitution instance of "means μ_a" only if "plane" expresses the idea of an airplane in that instance. The same goes for "means *plane*" and "means plane." "Means 'airplane'" is also a substitution instance of "means μ_a," assuming that "airplane" has its salient English sense. If *w* is the idea of a carpenter's wood plane, then "means 'plane'" is a substitution instance of "means μ_w" only if "plane" expresses the idea of a wood plane in that instance. "Means 'airplane'" could not be a substitution instance of "means μ_w" unless we were using a code of some sort. With these conventions in hand, we can use the following schema to define expressions of the form "the idea μ."

7.5 **Definition:** *The idea μ_i = i, provided i is an idea.*

When "μ" does not express an idea, "the idea μ" has no referent. Thus "the idea 'ouch'" has no referent because what "ouch" expresses is not an idea, and "the idea 'borogrove'" has no referent because "borogrove" does not express an idea or anything else. "Concept" is interchangeable with "idea" in Definition 7.5, and "c(μ_i)" is an abbreviation for "the concept μ_i." Hence c(μ_i) = i. Definition 7.5 tells us that "The idea *plane* is the idea of an airplane" is true when the occurrence of "*plane*" therein expresses the idea of an airplane. It would be false if "*plane*" there meant "wood plane." Together with Definition 7.4, Definition 7.5 entails:

7.6 **Theorem:** *e means μ_i (in L) iff e expresses the idea μ_i (in L), provided i is an idea.*

To say that "Flugzeug" means "plane," when we are using "plane" to express the idea of an airplane, is to say that "Flugzeug" expresses the idea of an airplane.

Definition 7.5 can be extended to cover specific *types* of idea as well. The *thought* μ_i = i, provided i is a thought. Hence the thought "The sky is blue" is the thought expressed by "The sky is blue," and an idea i is the *proposition* μ_i when i is a proposition (see Chapter 13). The generalization of Definition 7.5 would say that *the Ψ μ_i = i, provided i is a Ψ*. This holds when "Ψ" is replaced by "idea," "thought," "proposition," or any other general term denoting a type of idea.

Definition 7.5 applies in the same way to iterated idea descriptions. Thus *the idea "the idea 'vixen'"* refers to the idea expressed by *the idea "vixen."* Consider occasions on which I say that S used "the idea 'vixen'" to express the idea "the idea 'vixen.'" On some of these occasions, I used

"vixen" to mean "female fox," and I asserted that S used "the idea 'vixen'" to express the idea of the idea of a female fox. On other occasions, I used "vixen" to mean "shrewish woman," and I asserted that S used "the idea 'vixen'" to express the idea of the idea of a shrewish woman. In both cases, I would be implying that S *referred to* a first-order idea (the idea of a female fox or the idea of a shrewish woman) even though I claimed that S *expressed* a second-order idea.

Definition 7.5 tells us what "the idea μ" designates in the actual world. When generalized to all possible worlds, the definition would say that *the idea* μ_i = *i in any world w*. Expressions of the form "the idea μ" are *rigid designators*. On any occasion of use, they designate the same idea in all possible worlds. Whenever "the idea 'blue'" is used, it rigidly designates *that idea,* the one the word "blue" actually expresses on that occasion. If English were different, so that "blue" had a different meaning, then "blue" would express a different idea, an idea other than the idea of blue. The idea "blue" would remain the same, although a different phrase would designate it. It would remain the idea "blue" even though, in this hypothetical variant of English, it would not be designated by the phrase "the idea 'blue.'" This can be very confusing, for speakers of the imagined variant of English would say (and should say), "'Blue' expresses the idea 'blue,'" just as we do. The agreement would be verbal, however. Since their language would be different from ours, they would be talking about a different idea than we do when we use the sentence. Expressions like "the idea 'blue'" are thus indexical in the same way that "I" and "here" are. On different occasions of use, these expressions are used to rigidly designate different objects.

As noted, Definition 7.5 implies that sentences of form (1) are true in every context in which "the idea expressed by 'μ'" in its predicate is used to refer to the idea expressed by "μ" in its subject.

(1) The idea μ is the idea expressed by "μ."

It similarly follows that "The idea μ is the idea expressed by 'μ' *on this occasion*" is true whenever "this occasion" refers to the very occasion on which that sentence is being uttered. Definition 7.5 does not imply that "the idea μ" and "the idea expressed by 'μ'" are *synonymous*. They are not. First, when speakers use "the idea *blue*" in ordinary usage, they are *using* the word "blue" but they are not talking *about* that word. By contrast, speakers are talking about the word "blue" when they use "the idea expressed by '*blue.*'" Second, sentence (1) could be used to say something false. The speaker could be using the singular term in the predicate of (1)

to refer to the idea expressed by "μ" in some contextually salient place other than the subject of (1) itself. If "μ" is '*blue,*' for example, the speaker might be pointing at an occurrence in a book, where it meant "sad," even though the speaker used it to mean the color blue. The speaker could also use "the idea expressed by 'blue'" in its unrestricted Russellian sense. In that case, the instance of (1) is either false or neither true nor false due to failure of the uniqueness presupposition. Being ambiguous, "blue" expresses more than one idea. Third, substitution instances of (1) do not express necessary truths even when they are true. The idea *blue* need not be expressed by "*blue*" in English at all, and need not be the idea expressed by "*blue*" on any particular occasion. The singular term in the subject of (1) is a rigid designator, the singular term in the predicate is nonrigid. Substitution instances of (1) thus resemble "Here is the place where I am," which is true in every context in which "Here" and "I" are used deictically, but is not analytic and does not express a necessary truth in any context. Because (1) is about the word "μ," can be used to say something false, and never expresses a necessary truth, it cannot be listed as a theorem. Nonetheless, Definition 7.5 does account for why (1) is true on its most natural interpretation.

Definition 7.5 similarly implies that (2) is true on any occasion on which "w" is used with one of its conventional ideational meanings, and the speaker is talking about the language he is using.

(2) "w" expresses the idea "w."

Assume that i is the idea "w" expresses after "idea." Then "the idea 'w'" refers to i in (2), and (2) consequently asserts that "w" expresses i, which is true if, as assumed, "w" was used in (2) after "idea" with one of its conventional meanings. Given Definition 7.4, it follows further that every instance of (3) is true on any occasion on which "w" is used conventionally.

(3) "w" means "w."

Thus in the previous paragraph, "necessary" expressed the idea "necessary," "truth" meant "truth," and so on. Despite the tautological character of (2) and (3) on their most natural interpretations, they are not theorems, for the same reasons that (1) is not. If Sally mistakenly thinks "googol" means 10^{10}, then she could use "'googol' means 'googol'" to express a falsehood, namely, that "googol" means "10^{10}." If Sally is using American English to describe British English, and says "'hood' expresses the idea 'hood,'" then she has asserted something false.

Note well that quotation has different functions on the two sides of "means" in (3) or "expresses the idea" in (2). Only on the left does quotation form the name of the word quoted. On the right, the quotation marks emphasize that a word is being used ideo-reflexively. Hence "'rot' means 'red'" is equivalent to "*The word* 'rot' means 'red'," but not to "'rot' means *the word* 'red.'" It is not even true that *"rot" means "red"* is equivalent to *"rot" means the same as "red."* For a Chinese woman who knows no German or English may know that "rot" means the same as "red" without knowing that either "rot" or "red" means "red."[19] Furthermore, consider a possible world w in which "rot" means what it does in the actual world, but "red" has a different meaning. Then "'Rot' means 'red' in w" is true, but "'Rot' and 'red' have the same meaning in w" is false. On the left, "w" is not used with any of its conventional meanings. When a man says "'Rot' means 'red,'" he does not mean "red" or anything else by "rot." He only means "the word 'rot'" by "rot." By contrast, we need to know whether the man means "colored red" or "communist" by "red" in order to know what he is saying about "rot." If a woman says "'Ente' means 'duck,'" we need to know whether she means "bird of the family *Anatidae*" or "dip down in avoidance" by "duck" to know whether she is stating a truth or a falsehood.

Horwich (1998a: 16–17) would argue that the knowledge that "red" means "red" is trivial, something that we can work out from the convention whereby we quote or italicize a word after the word "means" in order to state what the word means. He draws the startling conclusion from this that understanding the word "red" cannot consist in knowing that it means "red." This argument embodies a use-mention confusion. Knowing that "p" is true is not the same as knowing that p. A Chinese woman who knows no English might come to know, because I gave her a list of true sentences to memorize, that the sentence "Washington is north of Richmond" is true. But she can memorize that list of sentences without knowing anything about Richmond, and hence without knowing that Washington is north of Richmond. Similarly, my young son John knows enough about the English language to know that sentences of the form "'w' means *w*" are true, from which he might infer that the sentence "'chien' means *chien*" must be true. He can know this without knowing that "chien" means *dog*. But to know that a word means *chien* is to know that it means *dog*. Consequently, most Frenchmen know that "chien"

19 Cf. Alston 1963a: 407, fn. 2; 1964a: 20–2; Lewy 1976: Chapter 1; McDowell 1977: 169; Sellars 1979: 91, 113; Kripke 1982: 9, fn. 8; and Heidelberger 1982: 23–4.

means *dog* without knowing what the English word "dog" means. To know what a word means is to know what idea it expresses. To know that is to know an important fact about how the word is used, not something trivial.

As Definition 7.5 reflects, expressions of the form "the idea μ" display one kind of compositionality but not another. The *meaning* of "the idea 'vixen'" is a function of the *meanings* of "the," "idea," and "vixen." In particular, changing the meaning of "vixen" from "female fox" to "shrewish woman" changes the meaning of "the idea 'vixen.'" "The idea '2'" and "the idea 'two'" have the same meaning because "2" and "two" do. The reference of "the idea μ" is consequently a function of the meanings of its components. But the *referent* of "the idea μ" is not a function of the *referents* of its components. "The idea 'square root of four'" and "the idea 'cube root of eight'" refer to different ideas even though "square root of four" and "cube root of eight" have the same extension. To uphold the referential compositionality of "the idea μ," we would have to insist that "μ" is functioning as a singular term referring to the idea it expresses, which is the well-known Fregean fall from semantic innocence (see §21.2).

Definition 7.5 defines the *oratio recta* way of referring to ideas by displaying the words that express them, which can be used with any type of linguistic expression. When an idea is expressed by words in certain grammatical categories, local conventions using *oratio obliqua* are more common. For example, let "ρ" stand for any pronoun-free declarative sentence expressing a proposition, which is defined in §13.3 as a thought capable of being believed or disbelieved. The idea "ρ" can be referred to as the idea *that* ρ.

7.7 **Definition:** *The idea that* ρ = *the idea "*ρ*."*

Thus the idea that man evolved from the apes is the idea "man evolved from the apes." Instances of "the idea 'ρ'" in turn are defined by Definition 7.5. It follows that the idea that ρ_i = i, when i is a proposition. Like "the idea 'ρ'," "the idea that ρ" is a rigid designator. "The idea that the sky is blue" designates the same proposition even in a world in which "the sky is blue" expresses a different proposition. As with Definition 7.5, Definition 7.7 can be generalized to define "the Ψ that p," where "Ψ" stands for "thought," "proposition," and other general terms denoting kinds of ideas. Thus the proposition that the sky is blue is the idea "The sky is blue."

The reason for restricting "ρ" to *pronoun-free* declarative sentences is that pronouns behave differently in *oratio recta* and *oratio obliqua*. There

are two differences. First, the reference of a pronoun is generally different. For example, "The idea 'I am a man' occurred to S" says that S was thinking about himself. But "The idea that I am a man occurred to S" says that S was thinking about me. It does not matter in either case who S is. Second, pronouns are partially transparent in that-clauses, completely opaque in the corresponding *oratio recta* form (see §6.2). "S is thinking the thought that I am a man" does not specify how S is thinking of me. That is, it does not specify the particular subject concept of the thought S is thinking; all it tells us is that the extension of the concept is me. "S is thinking the thought 'I am a man,'" by contrast, does specify how S is thinking of himself. It specifies that the subject concept of S's thought is specifically his self-concept.[20] The restriction on "ρ" is thus consistent with our general focus on the opaque interpretation of propositional attitude statements. It should be noted that even pronouns may not be fully transparent in that-clauses. "S is thinking the thought that I am a man" together with "I am the husband of Kathy Olesko" does not clearly entail "S is thinking the thought that the husband of Kathy Olesko is a man" (cf. §6.2).

We have observed on many occasions that applied to either words or speakers, "means" can take a that-clause as well as a quoted or italicized object. Definition 7.4 can now be supplemented by Definition 7.8.

7.8 **Definition:** *e means that ρ (in L) iff e expresses the idea that ρ (in L).*

Thus "The sky is blue" means that the sky is blue because it expresses the idea (thought, proposition) that the sky is blue. Definition 7.8 tells us that sentences of form (4) are true in any context in which the speaker is talking about English and is using "ρ" in one of its English senses after "that."

(4) "ρ" means that ρ.

Thus "'Grass is green' means that grass is green" is true as long as the speaker is talking about what "Grass is green" means in English, and is using "Grass is green" to mean something that it means in English. Sentence (4) is nevertheless not a theorem, for the sorts of reasons discussed in connection with (2) and (3) earlier.

That-clauses are ubiquitous, but familiarity does not always breed understanding.

20 See *Nondescriptive Meaning and Reference.*

It might reasonably occur to a philosopher to wonder, "Why is it that our canonical specifications of thoughts, beliefs and the like operate by employing phrases – embedded 'that' clauses – that (apparently) express actual or possible states of affairs? Why, for example, do we pick out the thought that it's raining by using the expression 'it's raining'? What is it about thoughts, and about states of affairs, that makes this practice possible?" . . . This is closely related to a revealing question that I believe was first raised by Donald Davidson: how are we to understand the fact that the expressions that can appear as freestanding declarative sentences can also appear as the complements of verbs of propositional attitude? (Fodor 1990a: 60–61)

Frege surely raised the question earlier, but his answer was flawed (§21.2), as was Davidson's (§13.4). The proper answer, I submit, is not at all profound. The facts about the thought that it is raining that enable us to refer to it using "it is raining" after "the thought that" are (1) the fact that "it is raining" expresses that thought in English; (2) if a thought is expressed in a context, it can be referred to; and (3) English contains the rule whereby the thought expressed by the sentence "ρ" can be referred to by "the thought that ρ." The fact that "it is raining" can be said to be about a state of affairs is inessential. Grammar requires that "that" be followed by a declarative sentence. Similar devices apply to thoughts expressed by other sentences that cannot be said to be about states of affairs.

We introduced the term *object nominals* to denote the class of nouns, noun phrases, gerunds, and nominalizations that can occur as direct or indirect objects (§6.1), and used "Φ" as a placeholder for them. Let us introduce "ϕ" as a placeholder for the more specific class of pronoun-free nominals. The idea "ϕ" can be referred to as the idea *of ϕ*. The idea of green is the idea "green," the idea of grass is the idea "grass," the idea of Mars is the idea "Mars," and so on.

7.9 **Definition:** *The idea of ϕ = the idea "ϕ."*

As in the case of that-clauses, pronouns behave differently in these two contexts. "The idea 'me' " is nonindexical, and designates the self-concept in every context. "The idea of me" is indexical, and in every context denotes some concept whose extension is the speaker. Note carefully that the idea "grass" is the idea *of grass*, not the idea *of* "*grass.*"[21] A Chinese person may have the idea of grass – which is the idea we English speakers customarily use "grass" to express – without having the idea of the English word "grass."

21 Contrast Alston 1964a: 24–5; Crimmins 1989: 291.

The placeholder "ϕ" must be interpreted as having narrow scope in Definition 7.9. When a phrase like "the idea of something" occurs in other contexts, however, the quantifier may also be interpreted as having wide scope, resulting in an ambiguity. For example, "If 'tuatara' is meaningful, then it expresses the idea of something," may be taken in two ways. If "something" has narrow scope, then Definition 7.9 applies, and the sentence says falsely that if "tuatara" is meaningful, then it expresses the idea expressed by "something." If "something" has wide scope, then the sentence says truly that "tuatara" is meaningful only if it expresses some idea. In general, when "something" has wide scope, "The idea of something is P" means that "The idea of ϕ is P" is true for some nonquantifier "ϕ."

Definition 7.9 is intended to define only the *opaque* sense of "the idea of ϕ" (§6.2). This is one reason why "ϕ" was restricted to exclude pronouns, because "the idea of Φ" is not fully opaque when "Φ" is a pronoun. "The idea of ϕ" almost invariably has the opaque sense when it occurs as the subject of a sentence (as in Definition 7.9), but often has the transparent sense when appearing in the predicate. Thus if "The idea of water is the idea of a liquid" were interpreted according to Definition 7.9, it would be falsely identifying the idea "water" with the idea "liquid." But the sentence would normally mean that the idea of water is the idea of something N, where N is a liquid, which is true because water is a liquid.

When "ϕ" is a simple gerund, "the idea of ϕ" is ambiguous in English, and has another meaning with an implicit personal pronoun that is more complex than the meaning defined by Definition 7.9. In this other sense, "the idea of V-ing" means "the idea of *one's* V-ing," which, in accordance with Definition 7.9, designates the idea expressed by "one's V-ing." Thus in "The idea of singing made Mary nervous," the subject denotes the idea of her (Mary's) singing. The same phrase has a simpler referent in "In 'Jane's singing made Mary nervous,' 'singing' expresses the idea of singing."

When the ideas we wish to refer to are expressed by words other than object nominals, the prepositional phrase construction cannot be used directly. Nevertheless, we often do use the construction with cognate nominals to produce a term whose reference is the idea expressed by the root of the nominal. When "A" is a suitable adjective, we often refer to the idea it expresses as the idea *of A-ness* or the idea *of A-ity*. (If "A" doubles as a noun, we can also refer to it as the idea *of A*.) We thus use "The idea of triangularity" to refer to the idea expressed by "triangular." When "V" is a verb, we often refer to the idea it expresses as the idea

of V-ing. The idea expressed by "walk" is called the idea of walking. When "p" is an indicative sentence, we may refer to the idea it expresses as the idea *of* NOM(*p*), where "NOM(p)" stands for the nominalization of p. For example, the idea "Mary sang" might be referred to as the idea of Mary's singing.

As a result, when "ϕ" is derived from a root form, "the idea of ϕ" is ambiguous in English. "The idea of Mary's singing" designates directly the idea "Mary's singing," which is different from the idea "Mary sang" that it designates indirectly. The latter is a thought, which may be true or false. The former idea represents an event, and is a subject-concept that appears in the thought that Mary's singing delighted the audience. Similarly, "the idea of triangularity" can be used to refer to the idea expressed by either the subject or the predicate of "Triangularity is not triangular."

Since the *oratio recta* form "the idea μ" is universally applicable and unambiguous, it is unnecessary for us to provide a definition that covers all instances of "the idea of ϕ" in English. The special case defined by Definition 7.9 is enough for all of our purposes, and together with Definition 7.7 suffices to indicate how the *oratio obliqua* form of ideo-reflexive reference works.

§7.7 WORD REFERENCE

Whether applied to words or to speakers, reference differs from meaning in three respects. One difference is that "refers to" must grammatically be followed by object nominals, which we are representing with the placeholders "Φ" and "ϕ." Another difference is that "refers" applies only to *verbal* (wordlike) expressions (see §6.1). A third difference is that word reference is transparent and fully relational, subject to both existential instantiation and the substitutivity of identity. "The forty-third president" refers to George W. Bush because he is the forty-third president. We do not know who, if anyone, "the seventieth president" refers to yet. And "the first female president of the twentieth century" does not refer to anyone. In this respect, word reference also differs markedly from speaker reference, which is opaque and intentional (§6.2). By making these modifications to Definition 7.4, we can say that *e refers to* Φ *iff (i)* Φ *exists and (ii) for some* $\Phi' = \Phi$, *e verbally expresses the idea of* Φ'. Thus "Roosevelt" refers to Franklin Delano Roosevelt in English because it is a word that expresses the idea of Franklin Delano Roosevelt, who was a real person. "The first female president of the twentieth century" does not refer to anyone, despite the fact that it expresses the idea of the first female

president of the twentieth century, because there was no such person. Using the convention discussed in §6.4 governing the use of quantification variables in normally opaque contexts, we can define word reference as follows:

7.10 **Definition:** *e refers to x (in L) iff e verbally expresses the idea of x (in L).*

Equivalently, we can say that e refers to x iff *e verbally expresses some i such that ex (i) = x*. The related notion of what a word refers to on an occasion of use will be defined in *Nondescriptive Meaning and Reference.*

We saw earlier that Definition 7.5 explains why "the idea 'ϕ' is the idea expressed by 'ϕ'" (1) is true in every context in which "the idea expressed by 'ϕ'" in its predicate is used to express the idea expressed by "ϕ" in its subject. Together with this, Definition 7.10 explains why instances of the "disquotation formula" (5) are true in any context in which the speaker is talking about English and is using "ϕ" in one of its conventional senses.

(5) "Φ" refers to Φ, provided Φ exists.

Thus we can truly say that "Venus" refers to Venus, "2" to 2, and so on. Sentence (5) is not a theorem, however, and can be used to assert a falsehood. Thus "'One billion' refers to one billion" is false if I am using British English (in which "billion" means "million million") to describe American English (in which "billion" means "thousand million"). This is a mistake I am sure at least some Britons have made when reading the American financial papers.

§7.8 USING A LANGUAGE

Languages are abstract objects. Once a set of expression-idea pairs is specified, it can be studied without reference to speakers. It may be immensely difficult to specify the grammatical and semantical rules of the language, but the task is purely formal given the expression-idea pairings. Other questions about languages, however, cannot be pursued in the abstract, such as these: What is a *natural* language? Which natural languages exist? How did they arise? How and why do they change? Which languages are spoken in a particular country, region, or community? What language is a particular speaker using? The last of these questions is the most fundamental.

Infinitely many languages exist as abstract objects. But only some are actually used by any given individual, and the total number of languages

used by intelligent beings to date is surely finite. To *use* a language, it is obviously necessary for the individual to use some of the expressions of the language. A speaker who does not utter a word of English is not using English.[22] He may understand English, and have the ability to use it; but he is not actually using it himself. A French-Canadian who knows English just as well as he knows French might refuse to speak English for political reasons. He can truly say (in French), "I do not use English." Finally, a person who is merely reading a book in French, without saying a word, is using his knowledge of French, but is not using French. If our French-Canadian overhears an English lady and understands her, he is not violating his vow never to use English.

Using words from a language is just as clearly not enough to count as using that language. First, combinations of the words used must also be part of the language. A person who says things like "Ball if saw and," "Oh an that heave collection" is using English words but not speaking English. Second, the speaker must use the words and constructions to express what they express in the language. That is, he must use "modes of expression" from the language. He must use some of the ways of expressing ideas that constitute the language. A cryptographer who said "The cat is on the mat," but meant "Enemy submarines have surfaced" used an English sentence but did not use English. American tenors who sing operas in Italian are not thereby using Italian. Only a fool would infer from a fine performance of a Verdi opera that the singers can speak Italian.

It is not necessary, of course, that S's expression-idea pairings *exhaust* L. Natural languages contain infinitely many phrases, clauses, and sentences, of which any given speaker uses only a finite sample. Indeed, a speaker might use only one word of the language. Thus S may use Spanish by answering "Dos" to the question "How many brothers do you have?" It may be thought obvious that using Spanish requires more than this, since what I did hardly proves that I *know* Spanish. But we are not trying to define that difficult notion. It is quite clear that S was *using* Spanish, and using it specifically to communicate that he has two brothers.

22 Lewis (1975: 12) seems to deny this. But he explicitly claims only that S may be a member of a language using community without (ever) using expressions of that language. This can be granted. The United States is an English-speaking nation, for example, even though not all U. S. citizens speak English, or are able to speak English. More pertinently, an American may know English and understand others who use it, but be unable to use English owing to general paralysis. In general, S's membership in a group that uses L does not entail that S uses L. And the fact that S uses English does not entail that S is currently using English. Lewis did not define what it is for an individual to use a language.

Earlier, I described English as the combination of all of its dialects. Ziff saw a problem in this maneuver that can arise only if we assume that using a language requires using all of it.

> If we say that 'I wull.' is an English utterance because it is an utterance of Dorset and 'Ahm goan.' (i.e. 'I'm going.') is an English utterance because it is an utterance of dialects in southeastern United States, we shall be describing a language that no one speaks: apart from special contexts, such a pair of utterances is not likely to be uttered by a single speaker. (Ziff 1960: 4)

It would of course be absurd to conclude that no one speaks English. But this conclusion does not follow from the equally undeniable fact that no speaker uses every word of English, nor from the likely fact that some pairs of English words have not been used by any single individual (consider "zymosthenic" and "abolla"). Simpson makes the same mistake in his refutation of the view that Dutch and German are different languages.

> Allusion was made above to the popular view of *a language* as 'a system of speaking or writing common to a group of people'; but in the case of each dialect continuum mentioned above no single system *is* discernible. (J. Simpson 1994: 1895)

What the German dialect continuum shows, for example, is that there is no system of modes of expression such that all German speakers use all of them. Speakers in Köln use *dorp*, speakers in Koblenz use *dorf*, but speakers in both places use German.

It is also unnecessary for the speaker to use words *exclusively* from the language he is using, or to use them exclusively to express what they mean in that language. Speakers of English frequently introduce new words (often borrowed from another language), or give existing English words new meanings. Until the words "catch on," they are not part of English. Bilingual speakers, or those still learning English, may dot their speech with words from another language, such as Spanish. Even competent users of a language occasionally misspeak, using the language incorrectly. So the most that can be required is that the speaker's expression-idea pairings be *predominantly* those of the language. This means that the concept of using a language has at least one element of vagueness. If every expression-idea pairing S used is from L, then S is clearly using L. If none of the expression-idea pairings S used is from L, then S is clearly not using L. In between, it may be unclear whether or not S is using L. The further from 100 percent and 0 percent the proportion of expression-idea pairings from L is, the less clear it will be. There is no precise percentage above

which L is clearly used, and below which it is clearly not used. There is also no way of weighting the relative importance of lexical, orthographic, grammatical, and semantic deviations. Other terms denoting complex human activities are similarly vague. For example, if S played every note of the *Moonlight* Sonata in the right order, then S played the *Moonlight* Sonata. If he played a few wrong notes, he still played the sonata. If he did not play a single note of the *Moonlight* Sonata, then he did not play that sonata. In between, matters are unclear. There is no precise percentage of wrong notes above which he clearly is not playing the sonata, and below which he clearly is.

Let us stipulate that S *conforms* to L when the expression-idea pairings he uses are at least predominantly from L, or in other words, when S generally follows the rules of L.

7.11 **Definition:** *S conforms to L iff S uses modes of expression predominantly from L.*

According to Definition 2.5, S meant μ by e iff S used e to directly express the idea μ. Hence S conforms to L iff S uses expressions and constructions predominantly from L, and uses them predominantly to mean cogitatively what they mean in L. Conforming to a language involves a coincidence of word meaning and *cogitative* rather than cognitive speaker meaning (see §2.4). Most of the expressions a speaker uses – such as interrogative and imperative sentences, and proper parts of declarative sentences – are not used with any cognitive speaker meaning. Moreover, when a speaker uses a declarative sentence figuratively, only cogitative speaker meaning will coincide with word meaning. In irony, for instance, cognitive speaker meaning is generally the opposite of what the sentence means.

Conforming to L, as we have defined it, is not sufficient for using L. In general, if S conforms to L, then S will automatically conform to an infinity of other languages: all those containing either L or the part of L that S actually used. But S will be using only a small number of those languages. Suppose S is an American speaking English who says "The sky is blue," meaning just that. Unbeknownst to S, an airline company has a secret "Code Delta" based on English. While Code Delta differs significantly from English, some of its expressions have the same meaning as in English. "The sky is blue" and its components are examples. Then every mode of expression S used is in Delta. But it is incorrect to say that S used that code. Similarly, S used only expressions that are common to both American and British English. But even though S is using expressions and constructions exclusively from British English, and using them to mean

what they mean in British English, it would be unusual if an American were using British English rather than American English.[23] Finally, S is conforming not only to English, but to the combination of English and Hopi. Yet it is unlikely that S is using any such system of expression.

What is missing in these cases is another layer of intentionality. While S conformed to Code Delta, British English, and the union of Hopi and English, he did not do so on purpose. S did use "The sky is blue" to express the thought that that sentence happens to express in Code Delta. But S's reason for using that sentence did not include the fact that it expresses what it does in that code. Similarly, S probably used the words he did because of what they mean in American English, not because of what they mean in British English. Finally, normal people never think of the union of English and Hopi, and would not try to conform to it even if they did. I believe that intentional conformity to a language does constitute using it.

7.12 **Definition:** *S uses L iff S intentionally conforms to L.*

In other words, S used L iff it is generally true that S used e to express i (or mean μ) *because* e expresses i (or means μ) in L. It follows that S cannot use a language without having *learned* or *constructed* it. For S would not then be able to form the intention to conform to it. It similarly follows that S *knows* the language, or at least *some of it*. Definition 7.12 is silent on the issue of whether S knows any of the rules of L other than the material rules. For it seems that applying the material rules is sufficient as well as necessary for using L.[24] Given the use of the individual variable "L" in the normally opaque context created by "intentionally," Definition 7.12 says that a speaker who is using the language with the largest vocabulary must be trying to conform to that language, but need not think of it as the language with the largest vocabulary (§6.4). He must think of the language under some description, but not under any particular description. Put another way, the speaker must have in mind a certain abstract structure pairing expressions and ideas, and must be using it as a model to determine

23 These examples undermine Lewis's (1969: 183) suggestion that S is using L as long as he has the right expectations and preferences "*in sensu diviso*" rather than "*in sensu composito*." Contrast also Blackburn 1984: 26. Cf. the distinction between "fitting" a rule and being "guided" by it, developed by Quine (1972: 442), Rosenberg (1974: 31), and others. See also Lewis 1975; 1992 and O'Leary-Hawthorne 1990; 1993.

24 Definition 7.12 may still be too weak. Now that we are thinking of the union of English and Hopi, suppose that we use "The sky is blue" to mean "The sky is blue," because it has that meaning in the union of English and Hopi. Are we using the union of those two languages? I am inclined to say yes, but do not have a firm opinion.

what he should say. But he need not have that structure in mind in any particular way.[25]

If S used L, then he used it *in a given place*, *on a given occasion*, *at a certain time*, *during a certain time interval*, and *from one time to another*. Definition 7.12 should be understood as implying that S used L *at t* iff S intentionally conformed to L *at t*, and so on for each adverbial of time or location. Any ambiguity in these adverbials will produce an ambiguity in the sentences containing them. "There," for example, may refer to a precise spot, or to the general neighborhood of that spot. Imagine that Bill puts his finger on a page and asks "What language was Steve using there?" And suppose that Bill put his finger on the one French word that Steve used, the rest of the words being English. Then the proper answer would be "French" if Bill meant that spot on the page, or "English" if Bill meant that whole page. Finally, the more specific notions of *speaking* or *writing* a language (or *in* a language) can be defined as intentionally conforming *in speech* or *in writing* to L. This distinction turns on how S produced the expressions of L he used.

The fact that people who are said to speak a common language may exhibit great variety in their speech has led some to the conclusion that languages as rule-governed systems are somehow unreal.

> The synchronic language-system is a theoretical construct of the linguist; and it rests upon the more or less deliberate, and to some extent arbitrary, discounting of variations in the language-behaviour of those who are held, pre-theoretically, to speak the same language. If pressed, we have to admit that there is a somewhat different language-system (a different idiolect*) underlying the language-behaviour of every individual, and that this too changes through time. (Lyons 1977: 243)

It is important to notice, therefore, that Definition 7.12 together with Definitions 7.11 and 7.1 allow that Americans and Britons both use English, even though the former but not the latter use American English. Simplification, approximation, and the like are as useful and inevitable in the description of widely spoken languages such as English as they are for any natural phenomenon of great complexity. But idealization is not committed until the theoretician presents a set of grammatical or semantic rules for English that holds only for "standard" dialects of English, failing for dialects different in minor ways from the standard. In other words, it is not the claim that people with very different individual speech patterns all use a common language such as English that is

25 Cf. Lewis 1969: 50.

false. What would be false is any claim that this common language obeys certain rules that fail for some of its dialects, or any claim that all speakers of the common language have internalized rules that fail for some of the individuals.

§7.9 APPLIED WORD MEANING

We now have the materials to analyze what I shall call *applied word meaning*, the notion of what an expression means *here*, or *on a given occasion*.

(6) e means μ on o.[26]

This is equivalent to:

(7) e means μ as used by the speaker on o.

Roughly put, the meaning of an expression on a given occasion is the sense that the speaker intended it to have on that occasion. More accurately, applied word meaning is the coincidence of linguistic word meaning and cogitative speaker meaning.

7.13 **Definition:** *e means μ on o iff (i) the speaker S means μ by e on o, and (ii) e means μ in the language S is using on o.*

Suppose that S is using English on the occasion we are interested in, and says "John bought a car," meaning "automobile" rather than "rail car" by "car." Later, S says "The locomotive is pulling thirty cars," meaning "rail car" rather than "automobile." The next day, S transmits "Five cars arrived" in a secret code in which "car" means "light cruiser," so that he means "Five light cruisers arrived." Then on the first occasion "car" meant "automobile," on the second occasion it meant "rail car," and on the third it meant "light cruiser."

Note that applied word meaning must be defined in terms of cogitative rather than cognitive speaker meaning. When Horowitz said of Rachmaninoff's Third Piano Concerto "It is an elephant," he meant cogitatively "It is a member of the family *Elephantidae*," while meaning cognitively that it is a huge and difficult piece of music. His words meant the former rather than the latter on that occasion. Recall also that the notion of language use employed in Definition 7.13 was itself defined in

26 Cf. Grice 1968; 1969; Kripke 1977; and Yu 1979: 272, 279. See also Katz and Fodor's (1963: 487–8) distinction between what words mean "in a setting" and what they mean "in isolation," and Bach and Harnish's (1979: 20–3) notion of "operative meaning."

terms of the coincidence of word meaning and cogitative speaker meaning (§7.8). The example also shows that what words mean on an occasion cannot be equated with what the speaker used them to *say* (contrast Davidson 1974: 456). For Horowitz did not say that Rachmaninoff's Third Concerto is a member of the family *Elephantidae*. Finally, it will typically be true that when e means μ on a given occasion, the speaker intended e to mean μ on that occasion (cf. Wilson & Sperber 1981: 156). But "S intended e to mean μ on o" cannot replace "S means μ by e on o" without rendering Definition 7.13 circular. Such a replacement would also make the definition too strong, for a word may well mean something on a given occasion even though the speaker is too young or too unreflective to have acquired the concept of applied word meaning.

We say that an expression is *ambiguous on a given occasion* when we cannot figure out which of several things the speaker means by it, or when the speaker engages in double entendre ("Frank's Nursery: A Growing Concern!"). An expression is *ambiguous in a language* when it means more than one thing in that language. Thus "car" is ambiguous in English, since it means both "automobile" and "rail car." It is conventionally used to mean both things, even though on any given occasion it is used to mean only one of those things. A word is said to have *no meaning* on a given occasion when the word means nothing in the language S is using (he utters a nonsense word), or when the speaker means nothing by it (he is just testing his voice, or practicing his penmanship). An expression also fails to have an assignable meaning on a given occasion when the speaker makes a mistake and means something by it that the word does not mean in the language he is using. Suppose S says "That is a vixen," meaning "baby cow" by "vixen," and believing erroneously that he is speaking proper English. Then we cannot say that "vixen" meant "baby cow" on that occasion, nor that it meant "female fox," "shrewish woman," or anything else. We cannot say what the word meant on that occasion.[27]

"Occasion meaning" would have been an excellent name for applied word meaning, except for the fact that Grice (1968; 1969) used that term to denote speaker meaning, which also varies from occasion to occasion.

27 Nor, for reasons I cannot divine, would we say that the word meant nothing on that occasion. Taschek's (1995a: 289–93) Paderewski case (S believes that there are two Paderewskis, one a pianist and the other a statesman, and sometimes uses the name to refer to one, sometimes the other) is another in which S means something definite, the name has a definite meaning in the language, and yet the name used has no assignable meaning on occasions when S uses it.

Grice used "timeless meaning" to denote linguistic word meaning, thus opposing it to both "occasion meaning" (speaker meaning) and "applied timeless meaning" (applied word meaning). I have avoided "timeless" because it is misleading. Word meaning may appear to be timeless *compared* to speaker meaning and applied word meaning, since the latter often vary considerably within a short span of time while the former changes slowly (by the standards of an individual speaker). But linguistic word meaning does change over time, and is expressed by a tensed verb. As noted earlier, "jument" *meant* "pack horse" in French a long time ago, but today *means* "female horse"; it is impossible to predict what if anything it *will mean* a thousand years from now.

Grammatically, the distinction between (8) and (9)

(8) e means μ. (*Linguistic Word Meaning*)
(9) e means μ on o. (*Applied Word Meaning*)

is like that between (10) and (11):

(10) S means μ by e. (*Customary Speaker Meaning*)
(11) S means μ by e on o. (*Occasional Speaker Meaning*)

But semantically, the distinctions are different. "On o" functions just like an adverb of time in sentence (11), and (10) is simply the "frequentive" mode of (11) (when the former is not just short for the latter). That is, sentence (11) is true just in case S used e to mean μ *at* t_o, where t_o is the time of o. Sentence (9), by contrast, does not mean that e meant μ *at* t_o. Let us suppose that the occasion in question was a moment twenty years ago, and that e is "vixen." Then, as now, "vixen" meant "female fox" in English. It also meant "shrewish woman" at that time. But on that occasion, it is highly unlikely that "vixen" meant "shrewish woman" if it meant "female fox" on the same occasion. In other words, we can treat the difference between (10) and (11) as simply a difference in tense, mode, or "form" of the word "means." Each form is defined by a variant of Definition 2.5 by letting the main verb in the definiens have the same form. But the difference between (8) and (9) (when the former is not simply short for the latter) is more substantive. The definition of (9) cannot be obtained from that of (8) by any simple variation of the definiens.

It should be noted that in sentences (6) through (11), as well as in Definition 7.13, the variable *e* ranges over expression *types*. Many philosophers have spoken of the meaning of a word *token*, or word

utterance.[28] This usage is unconventional: "mean" does not apply to tokens or utterances in any of its standard English senses. I have never seen an explicit definition of token meaning, furthermore, and the intended meaning is not entirely clear. I believe that "the meaning of a word token" is intended to denote what the speaker meant by the word type it tokens. But it might instead denote what the word type means on that occasion. The two definitions generally coincide, but are not equivalent, as we have seen. If Sam slips and says "erotic" instead of "erratic," then the token of "erotic" that he produced meant "erratic" according to the first definition but not the second. For the speaker meant "erratic" when he said "erotic." But the word "erotic" does not mean "erratic," even on this occasion. Similarly, children regularly describe certain department store employees as "Santa Claus." The children were referring to those employees. But the name "Santa Claus" does not refer to them, even on those occasions. Since the application of "means" to tokens or utterances is at best an unconventional way of expressing a type of meaning for which conventional means of expression are available, we will not adopt this usage.

28 Cf. Grice 1957: 385; Patton & Stampe 1969: 8; Schiffer 1972: 159; Field 1972: 274; Chastain 1975; Katz & Langendoen 1976; Katz 1977b: 35; 1994; Kasher 1982: 26; Pettit 1987: 727–8; Blakemore 1992: 5; Patton 1994: 44; Braun 1995a; Wreen 1998: 321. Others have spoken in a similar way of what a word means "on the lips" or "in the mouth" of a given speaker; see Evans & McDowell 1976: xx; Pettit 1987: 727–8; Peacocke 1986: 115–16; Stanley 1997a: 564. Travis (1997) muddies the waters with "speakings of words."

8

Basic Word Meaning

Chapter 2 began with the fundamental distinction between speaker meaning and word meaning. Even though "banana" means a kind of fruit, Alfred Kahn once used it to mean a decline in the gross national product in order to make fun of politicians who wanted him to avoid the terms "recession" and "depression" because of their negative associations. We have defined what it is for a speaker to mean or express something in terms of the speaker's intentions. We must now begin analyzing what it is for a word or other expression to mean something or to express an idea. What makes it true, for example, that "vixen" means "female fox" in English, but does not mean anything in French? Why does "jument" mean "female horse" in French today, whereas earlier it meant "pack horse"?

It seems self-evident that such facts about English and French are determined in some way by facts about the individuals who have actually spoken English and French. It would be as hard to maintain that an abstract system such as English arose and evolved independent of the actions of concrete human beings as it would be to maintain that the species *Canis familiaris* arose and evolved independent of concrete dogs. Most, if not all, facts about natural languages are determined by facts about their users, such as: whether they are spoken by millions of people or just in a particular geographic area; whether they are spreading or dying out; whether they have a common origin; whether they were influenced by other languages, or isolated; whether they are commercially or politically important; whether they are taught in various schools; whether they are written as well as spoken; whether they have many dialects or only a few; and so on.

It is not at all obvious, however, just which facts about individuals determine what words mean in a language. Grice proposed that word

meaning is determined by cognitive speaker meaning. We will see in this chapter that although Grice's own line of analysis is more promising than associational or truth conditional analyses, it has a number of limitations. These can be overcome by replacing cognitive speaker meaning with cogitative, and introducing the notion of convention. When cogitative speaker meaning is understood in terms of the expression of thoughts and ideas, the result is a more Aristotelian or Lockean analysis of word meaning. But since the analysis is still in line with the general Gricean program, I will call it the neo-Gricean analysis. We will observe that conventions to use words to express or indicate certain ideas are correlative with conventions to use them to communicate those ideas and with conventions to interpret the words as expressing the ideas.

Following customary usage, we have called the type of meaning we are now interested in "word meaning," in order to contrast it with speaker meaning and evidential meaning. But we must bear in mind that what we are calling word meaning is possessed not just by individual words, but by linguistic expressions generally, including sentences, clauses, phrases, and many word segments. Linguists use the term "morphemes" for the "minimal meaningful units of the language," and analyze the word "hunters" into the morphemes "hunt," "er," and "s." Morphemes are meaningful in the same sense that the words composing them are, so morphemes too have word meaning. We will extend the neo-Gricean analysis beyond the basic case of lexical word meaning to compositional word meaning in Chapter 10. Having taken cognitive speaker meaning as fundamental, Grice's analysis was ill-suited to handle subsentential meaning. For only sentences are used to mean that something is the case. We shall argue at length that the sentential primacy thesis, according to which subsentential word meaning can be defined in terms of sentence meaning, is untenable. The neo-Gricean analysis has no difficulty with subsentential word meaning, since words as well as sentences are used to express ideas, and speakers therefore mean things by both.

It is also unimportant for our purposes exactly how words are defined or individuated.[1] Sometimes we say that English contains different words that are spelled and pronounced the same ("homonyms"), an example being the verb "fall" and the noun "fall." On other occasions, we say that the word "fall" is ambiguous, meaning "descend due to gravity," "descent

1 Cf. Quine 1960: 129–30; Alston 1964a: 60; McCawley 1968: 125–6; Grice 1968: 132–7; Lyons 1971: 68–70; Kripke 1980: 7–8; Kručera 1982: 39; Evans 1982: 68–9; Jackendoff 1983: 111; Blackburn 1984: 20–1; Kaplan 1990b.

due to gravity," or "the season after summer." Sometimes we say that the word is pronounced /fawl/ and spelled f-a-l-l. But at other times we distinguish between the written word "fall" and the spoken word "fall." "Dreamed" and "dreamt" may similarly be counted as variant spellings of the same verb, or as different words with the same meaning. "Playing" and "played" may be considered different words with the same root, or different forms of the same word. No matter what definition or principle of individuation we adopt, words have meaning in the sense that we are interested in. Words in any sense are intrinsically meaningless. There is something about language users that makes it true that words have the meanings they do. We want to know what that is.

§8.1 ASSOCIATIONAL ANALYSES

Arnauld offered what was perhaps the first explicit attempt to define what words mean in the symbolic rather than the evidential sense.

> To say that a written or spoken word means such and such is to say only that our minds entertain the meaning, that is, the idea connected with that word whenever we hear or see the word. (Arnauld 1662: 90)

Arnauld was proposing that e means μ provided that e is *associated with* the idea μ.[2] As he immediately noted, however, a word that means μ is associated with many ideas other than the idea μ. For example, hearing the word "salt" makes the ideas of pepper, spice, food, sodium chloride, water, Morton, iodine, and many other things occur to English speakers. Arnauld subsequently equated the meaning with the "principal" idea connected with a word, but gave no criteria for separating the principal idea from accessory ideas, a defect shared by more behavioral versions of the theory.[3] It is equally clear, moreover, that none of us knows the meaning of every word in our language, and that we often forget those we do. So even the "principal" idea does not occur to us *every time* we hear or see the word. Finally, an expression is sometimes associated with different ideas in different countries. So there may be several "principal" ideas.

2 See also J. Mill 1829: 388–9; and Chapter 18 of this volume. Compare Ockham's "psychological notion" of a sign, according to which a sign signifies the thing it "brings to mind" (*Summa Logicae I*: §1; see also Loux 1974: 2; Freddoso 1980: 3). Carruthers' (1989: 101) nonideational theory of speaker meaning suffers from the same flaw.

3 Cf. Bloomfield 1933: 139; Morris 1946: Chapter 1; and Skinner 1957. Contrast Alston 1964a: 25–31 and Chomsky 1959.

The same difficulties can be found in some of the most recent accounts. The following was inspired by Chomsky, who postulated an innate special-purpose language processor.

> Expression meaning is simply that meaning which is associated with the utterance in the course of language processing directly through principles of the compositional semantics embodied in the language processor. (Laurence 1996: 295; also 282–3)

The problem with "salt" and "iodine" is avoided because it is not a semantic principle that associates the two. But this reference to semantics in the definiens obviously makes Laurence's definition circular. Without that reference, the analysis founders. It is presumably not the language processor alone that associates the idea "iodine" with the word "salt." But the language processor surely does associate the ideas "fox" and "female" with the word "vixen," even though the word means neither "fox" nor "female." Moreover, the language processor seldom associates an utterance of "broom" with the meaning *Cytisus scoparius* (even though the word has that meaning) and occasionally associates "entomologist" with the meaning "etymologist" (even though the word lacks that meaning). Furthermore, the language processor sometimes associates a word with different meanings in different countries. Finally, meanings are associated with few utterances in the early stages of development of the language processor. And no individual's language processor ever associates all meanings with all of the words in a natural language.

In a Hobbesian move, Horwich proposes to define what a word means in terms of the concept that its occurrence *indicates*.

> [W]hen we specify the meaning of a word we are also making a claim about what is to be inferred from its occurrence: we are specifying what a person who uses the word should be expected to 'have in mind'. Thus the intuitive construal of
>
> "perro" means DOG
>
> is something along the lines of
>
> The utterance of "perro" indicates (i.e. justifies belief in) the presence (within some mental state of the speaker) of the concept, DOG. (Horwich 1998a: 19–20)

While Horwich's general approach to meaning is quite congenial with what we will work out at length below, his particular formulation is much too broad. If the utterance of "perro" indicates the occurrence of any concept in the mind of the speaker, it surely indicates all of the

following concepts: *the word "perro," word, utterance, meaning*, *now*, *here*, *myself,* etc. From the fact that a native speaker uttered the word "perro," we can infer with great confidence that he had beliefs about words and what they mean – the word "perro" in particular – and that he realizes that he himself is uttering something here and now. It thus follows from Horwich's account that "perro" means *the word "perro," word, utterance,* etc. But it means none of these things. Horwich's definition is also too strong. As is well known, "bank" is highly ambiguous. One of the things it means is "billiard table cushion." Yet the other meanings of the word are so much more common that from the utterance of the word "bank" we can generally not infer the presence in the speaker's mind of the concept "billiard table cushion." The relative frequency of that concept's occurring given an utterance of the word "bank" is much lower than the relative frequency of other concepts' occurring. Utterances of the word "bank" do indicate occurrences of the concept "billiard table cushion" in certain contexts, to be sure. But due to the possibility of speakers' meaning something other than what the word means, utterances of the word "bank" can indicate the occurrence of any concept in certain contexts.

§8.2 THE GRICEAN ANALYSIS

Grice (1957; 1968; 1982; 1989) proposed that word meaning is determined by speaker meaning. This is plausible, since the two senses of the word "mean" seem to be closely related. There is clearly a close relationship between what words mean in English and what speakers of English mean by those words. It would seem completely self-contradictory, for example, to claim that the word "bachelor" means "unmarried man" even though *no one* had *ever* meant the latter by the former. But it is not nearly as obvious that word meaning depends on speaker meaning, or that the former can be defined in terms of the latter. Nor is it easy to say exactly how the two types of meaning are related.

Grice's original formulations were vague, tentative, and exploratory – not to mention crabbed. He proposed that "e means that p" is equivalent to some statement "about what 'people' (vague) intend (with qualifications about 'recognition') to effect by" e (1957: 385). Ziff (1967) responded that the effects that people intend to produce by uttering a sentence are determined by various factors, many of which have nothing to do with meaning, such as length and acoustic shape. Chomsky (1975: 63–4) and Yu (1979: 283–5) similarly objected that words may retain their meaning even though speakers do not intend to affect an audience by using

them. These problems with Grice's analysis of *word* meaning derived from defects in his analysis of *speaker* meaning, however. So they can be avoided easily enough by using the analysandum itself in the definiens rather than the analysans.

(1) Expression e means that p iff people use e to mean that p.[4]

Given the definitions of speaker meaning and expression provided earlier, this says that e means that p iff people use e to express the belief that p, that is, to produce thereby an indication of the belief that p, without pretense.[5] I shall call (1) the *Gricean analysis*. Grice himself suggested more complex analyses in later works (Grice 1968), as we shall later.

The observation that the meaning of an expression is determined by the way it is used was championed by followers of Wittgenstein (1953) and Austin (1962) under the slogan Meaning Is Use. This slogan is generally taken to imply a rejection of the theory that meaning consists in the expression of beliefs, thoughts, or ideas. For traditional use theorists held that meaning is determined by the use of words to do other things. One view focuses on the use of words to perform illocutionary acts like asserting, warning, requesting, ordering, and promising, which is undeniably one of the central functions of language. Hoping to avoid the regress and circularity objections discussed in Chapter 22, the most radical theories along these lines have attempted to define meaning without referring to the contents of the illocutionary acts. One proposal identifies meaning with *total illocutionary use*.[6] That is, two expressions have the same meaning provided that they have the same total illocutionary use.

(2) e means that p iff e is used to perform the same illocutionary acts that "p" is used to perform.

4 Cf. Chisholm 1958: 239; Stampe 1968: 165–74; Patten & Stampe 1969: 7; Schiffer 1972: 7; 1982: 123; Bennett 1976: 8; and Evans 1973: 300. See also Blackburn (1984: 118, 123–5), who, unfortunately, seems not to have fully appreciated the distinction between word and speaker meaning, and so would, I believe, regard (1) as circular and therefore uninformative.

5 If the analysans in the definition of expression (Definition 3.6) ends up embedded in a belief or other propositional attitude clause, and if the analysis is not analytic (that is, if the analysans is equivalent to but not synonymous with the analysandum), then these two alternatives may not be equivalent. Schiffer (1987a: 257–8) would object in that case that it would no longer be possible to claim that the semantic had been reduced to the psychological. But it would: for the lack of equivalence would be created by embedding speaker meaning in a *propositional attitude* clause, which means that word meaning would be reduced to a propositional attitude. As far as I can see, however, speaker meaning will be embedded in no such clause here.

6 Cf. Ryle 1957: 248–9 and Alston 1963a; 1964a: Chapter 2.

Unfortunately, (2) runs afoul of the elementary fact that sentences are usually ambiguous. For example, "Bill has no kitty" means "Bill has no young cat." But these two sentences do not have the same total illocutionary use. For the former but not the latter can be used to assert (roughly) that Bill has no small fund of money held in reserve, and the latter but not the former can be used to assert that Bill has no young member of the family *Felidae* (which includes tigers as well as domestic cats).

Instead of requiring that e be used to perform *all* of the illocutionary acts that "p" is used to perform, we might try requiring that e be used to perform only *some* of those acts. But then the theory becomes much too weak. For example, "Bill has an old cat" and "Bill has a young cat" can both be used to imply that Bill has a cat. But the two sentences do not have the same meaning. Another solution is to indicate in some way which meaning of "p" we have in mind, requiring that e be used to perform all of the acts that "p" can be used to perform *in that sense*. But a clause specifying what "p" means would make (2) circular or regressive as an analysis of meaning.

Another deficiency of (2) is that it is not evident what parallel account could be given for speaker meaning. Speaker meaning and word meaning are logically independent, as we have argued. But they are closely related phenomena nonetheless. An adequate theory of meaning should show that they are closely related. It is hard to see how (2) could meet this condition, since a speaker will seldom if ever perform *all* of the acts a sentence can be used to perform. For example, "You have an appointment today" can be used to inform Jane that she has an appointment, or merely to remind her. But it cannot be used on any one occasion to do both. If we try saying that S means that p by e if S used e to perform *some* of the illocutionary acts that "p" is used to perform, the result will be much too weak. It would follow, for example, that if S issued a reminder, then he meant "Did you remember your appointment?" as well as "Don't forget your appointment" by "You have an appointment."

These problems can be avoided by defining the meaning of an expression in terms of a *particular* illocutionary act, with content specified. An example of this sort of theory would be

(3) e means that p iff e is used to say that p.[7]

7 This is suggested by passages in Alston 1971: 35–6; 1974: 35–40; 1977: 17, 21, 23; 1980: 124, 127–8; 1987: 151, 157–8; 1994: 29, 36–8; Vendler 1972; Davidson 1974: 456; Dummett 1976: 72, 74; McDowell 1977: 160; 1978: 306–7; 1980: 121; Lycan 1984: 9; Skulsky 1986: 593; Schiffer 1987a: 1–6, 209, 123, 265; and Vanderveken 1990: 8. Alston focuses on what e

According to (3), a sentence is ambiguous if it is used to say more than one thing. What the sentence means on each interpretation is identified with what it can be used to say on that interpretation. Thus "A vixen is a mammal" means "A female fox is a mammal" because it is used to say that a female fox is a mammal. The same sentence also means "A shrewish woman is a mammal" because it is also used to say that.

One drawback of (3) is that it cannot provide much illumination about what it is for an expression to mean something until we have explained what it is for a person to say something. When we begin to analyze saying, it becomes clear that speaker meaning and word meaning are the more fundamental notions (see §5.3). If I turned on my turn signal, I may have meant, and communicated the information, that I am going to turn. But I did not *say* that I am going to turn. Similarly, if I described Joe metaphorically as a lemming, I may have meant that he is someone who follows the crowd no matter what. But I did not say that he is someone who follows the crowd. Saying, unlike meaning, requires the use of words that mean what the speaker means (with adjustments for indexicals and the like). But this reveals that (3) is implicitly circular (saying that p involves using an expression which means that p, which is what we are trying to define) and too narrow (a turn signal means something but is not used to say anything).

These defects of (3) can be avoided by selecting a more general illocutionary act. For example, *communicating the belief that p* does not require the use of meaningful expressions, and it does apply to turn signals. However, S may mean that p without communicating the belief that p to anyone. So an illocutionary use theory focusing on communication would be inferior to the Gricean analysis in ruling that a sentence used exclusively for recording secrets in a private diary has no meaning. Such a theory would make it not just practically but logically impossible for Robinson Crusoe to evolve his own distinctive idiolect. *Expressing the belief that p* solves all of the problems mentioned. The Gricean analysis may thus be viewed as a particular version of the use theory, in which the illocutionary act selected as constitutive of word meaning is the expression of belief. The superiority of the Gricean analysis to other illocutionary use theories should not be surprising, given that expression is the most general and

is *usable* to say. This change would improve (3) in some ways (never-used sentences), worsen it in others (figurative speech, indexicals). See also Grice 1968: 127 and Neale 1992: 553 in light of Grice's view that S means that p iff S either says or implicates that p (Neale 1992: 520).

fundamental member of the class that includes saying, asserting, warning, ordering, questioning, and promising.

Many have maintained that the Gricean analysis of word meaning is circular, either because speaker meaning must be defined in terms of word meaning, or because propositional attitudes like belief and intention themselves have content. I have already addressed the former suggestion (§2.1 and §4.5), and will take up the latter in Part IV.

§8.3 THE TRUTH CONDITIONAL ANALYSIS

Perhaps the leading alternative to Grice's proposal, and to mentalistic theories in general, is that the meaning of a sentence consists in its *truth conditions*.[8] A well-known difficulty with simple forms of the truth conditionalist theory is that having the same truth conditions does not guarantee having the same meaning (*Frege's Problem*). The sentences "8 divided by 2 equals the square root of 16" and "8 times ½ equals the cube root of 64" have the same truth conditions, but different meanings. Another objection is that interrogative and imperative sentences have meaning but no truth conditions (*Russell's Problem*). A deeper problem arises when we ask what it is for a sentence to have certain truth conditions. The usual answer is that to have certain truth conditions is to be true iff those conditions obtain. But sentences cannot be described as true or false absolutely. They may be true *in one sense*, false in another. "Grass is green" is true in one sense if and only if lawn grass is green. It is true in another sense if and only if marijuana is green. It is *because* the sentence has those senses that it has those truth conditions. What then are these senses relative to which, and because of which, sentences are true or false under certain conditions? That, of course, is just to ask what meaning is, which is the question the truth conditional theory was trying to answer.

Even if sentences had only a single set of truth conditions, the theory would leave a fundamental question unanswered, namely: *what determines*

8 See, e.g., Peirce 1931–5: §§5.476, 5.411–34; Carnap 1947; Quine 1953: 42; Davidson 1967; Kretzmann 1967: 395; Lewis 1969: 198; 1972; 1975: 20; Wiggins 1971: 17–18; Dummett 1973: Chapter 1; 1975; Wallace 1977; Kaplan 1977; 1989; 1990a; Evans 1982: 102; 1985: 72–3; Barwise & Parry 1983; Devitt 1981; Devitt & Sterelny 1987; Salmon 1986; 1989a; 1989b; 1990; Wettstein 1991. Contrast Fodor & Lepore 1992: 40–8; 216–18. The "information theoretic" approach sketched by Kittay (1987: 127) has most of the same defects, as does the verificationist (or falsificationalist) theory suggested by Dummett (1976) and the inferentialist approach developed by Block (1993: 1), Boghossian (1993), and Lance & O'Leary Hawthorne (1997: 8–10, 370). See also §23.3.

the truth conditions of sentences? That is, how do sentences come by their truth conditions? What makes "The Sun is massive" a sentence that is true if and only if the Sun is massive? What is it about that particular sequence of words that gives it truth conditions unlike practically every other sentence in the English language? Whatever it is, it is not intrinsic to the sentence. For the sentence could have had different truth conditions. Under what conditions would it have had different truth conditions? As I see it, these are among the most fundamental questions concerning the nature of language. It is not evident how the truth conditional theorist can answer such questions without bringing in the sorts of socio-psychological factors emphasized by Grice.[9]

To answer the fundamental question, Alston (1971: 35–6) appealed to the notion of *rules*, by which he meant conventional norms in force in the community. He came to say that for a sentence to have a certain illocutionary act potential is for it to be governed by a rule enjoining speakers not to use it unless certain conditions are satisfied. In the case of an indicative sentence, the conditions are its truth conditions. For a sentence e to mean that p is for there to be a rule making it permissible for speakers to utter e iff p.[10] Alston held that subjecting one's utterance of e to this rule constitutes asserting that p, and so derived as a theorem the principle that e means that p iff e can be used to assert that p (cf. (3) in §8.2). The conditions that need to be satisfied for permissible use of imperatives and interrogatives would be different.

Alston's theory is still subject to the intensionality problem. If it is impermissible to utter e unless A is a three-sided polygon, it is equally impermissible to utter it unless A is a three-angled polygon. Yet "A is a three-sided polygon" does not have the same meaning as "A is a three-angled polygon." Moreover, Alston needs to distinguish "semantic" impermissibility from other sorts that are irrelevant. There is a rule in our culture, for example, that one is not to speak with one's mouth full. Yet "The speaker's mouth is not full" is not part of the meaning or truth conditions of every sentence. The reason, obviously, is that the rule in

9 Cf. Searle 1983: Chapter 6; 1986, discussed in §4.5; and Lewis 1969; 1975, discussed in §11.8. See also Chierchia & McConnell-Ginet 1990: 81 and Horwich 1998a: 72. The truth conditionalist could of course take the Davidsonian line that the truth conditions of a sentence are determined by the *denotations* of its component words. But that just postpones the ultimate question. For now we must ask what determines the denotations of words. Why does "green" denote the set of green things?

10 See also Alston 1974; 1977; 1980; 1994; and compare the discussion of Alston's theory of expression in §4.5.

question is not a semantic rule. However, to incorporate such a restriction into Alston's analysis would render it circular.[11]

Like truth conditional theories in general, Alston's theory has a problem with ambiguity. "The grass is green" can be permissibly used provided that either the lawn grass or the marijuana is green. Yet "The grass is green" does not mean "The lawn grass or the marijuana is green." The sentence is ambiguous, not disjunctive. What we need to say is that *in one sense* "grass is green" is permissibly used if lawn grass is green, while *in another sense* it is permissibly used if marijuana is green. There are two different rules, one for each sense of the sentence. For a second reason, then, Alston's analysis is circular.

A third circularity resulted from Alston's handling of nonliteral speech.[12] "The president carries a big stick" may be permissibly used even when the president is carrying no sticks at all, or just tiny little sticks. For the sentence may be used metaphorically or ironically. Alston's solution was to restrict the governing rule to what is permissible in *literal* speech. But "literal" is a semantic term. Literal speech is that in which cognitive speaker meaning coincides with sentence meaning. Thus "The president carries a big stick" is used literally only if the speaker uses it to mean what the sentence means, namely, that the president carries a big elongated piece of wood. Indirect speech acts present a similar problem.

The solutions to the first and third problems jointly create another difficulty that is worse than circularity. To use a sentence meaning that p to say falsely that p is to lie, misinform, or mislead. These are not *linguistic* mistakes. The rule making it impermissible to use e literally to say something false is not a semantic rule. Hence there will not exist a rule meeting Alston's specifications for assigning the meaning or truth conditions to an indicative sentence.

Alston became enmeshed in all of these difficulties because he mistakenly believed that sentence meaning is determined by the truth conditional permissibility rules applying to them. On the contrary, which rules of this sort apply depends on the meaning of the sentence. There are conventions or rules connecting truth conditions to sentences. But

11 Alston (1964b: §4) himself saw no solution to this problem. A further difficulty is how Alston can avoid the result that "Jane has withdrawn" has as part of its meaning "The speaker believes that Jane has withdrawn," since it is normally impermissible for a speaker to utter the sentence without that belief, as Alston (1965: 26) observed in accounting for the concept of speaker expression (see §4.4, this volume).

12 Alston 1977: 30; 1994: 46–7.

they do so *indirectly*. The semantic conventions attach truth conditions to sentences *by* attaching senses to sentences (§23.3).

Horwich (1998a: 45–6) has recently sketched a variation of Alston's theory that replaces normative rules with descriptive rules governing acceptance. On this variant, the meaning of a word is constituted by the fact that it has a certain "acceptance property." The acceptance property is a disposition to accept certain sentences under specified circumstances that accounts for the acceptance of other sentences containing the word. For example,

> [T]he explanatorily fundamental acceptance property underlying our use of "red" is (roughly) the disposition to apply "red" to an observed surface when and only when it is clearly red. (Horwich 1998a: 45)[13]

Horwich suggests that the word "red" means "red" because this disposition involving "*a* is red" (where "*a*" designates an observed surface) explains our acceptance of all other sentences containing "red." Horwich does not show how this disposition explains the acceptance of other sentences containing "red." He does not, for example, provide any evidence against the possibility that our disposition to accept "*a* is red" is at least partly explained by our disposition to accept "*a* looks red" when *a* looks red, or by our unconditional disposition to accept "red is a color." Nor does he consider the possibility that all three dispositions are the result of a common underlying cause.

Horwich's analysis appears to suffer from the same intensionality problem as Alston's. I presume that subjects with the acceptance property just described would also be disposed to apply "red" to a surface iff they believed it to be a surface with enough red parts, or a surface that is shown to be red by the evidence. So why would Horwich's analysis rule that "red" means "red" rather than "surface most of whose parts are uniformly red" or "surface shown to be red by the evidence"? Horwich's analysis also suffers from the ambiguity problem. Whether we are disposed to accept "*a* is red" when "*a*" refers to a (believed) red surface depends on whether we take the word "red" to mean "red" rather than "communist" or something else. This move therefore introduces circularity into the analysis of word meaning. Furthermore, the fact that we are disposed to accept a sentence "*a* is *F*" iff *a* is a red object when we believe "F" to mean "red" tells us nothing about what "F" actually means. We would have this disposition

13 See also Horwich 1998a: 129, where names are the focus.

even if "F" actually meant "blue."[14] Horwich's (1998a: §3.11) suggestion that the word type "red" would have several phonetically and epistemologically indistinguishable subtypes with different acceptance conditions is no help. Let "red_r" be the subtype that unambiguously means "red" (the color). It remains true that we will accept "*a* is red_r" when *a* is red only if we believe that "red_r" means "red."

The Gricean analysis has a noncircular answer to the fundamental question that avoids the intensionality and ambiguity problems: the truth conditions of a sentence are determined by its meaning, which is determined by what people use it to mean. "The Sun is massive" is true iff the Sun is massive *because* "The Sun is massive" *means* that the Sun is massive. "The Sun is massive" has that meaning because people *use it to mean* that the Sun is massive. "Grass is green" is ambiguous because people sometimes use it to mean "lawn grass is green" and at other times use it to mean "marijuana is green." "A is a three-sided polygon" does not mean "A is a three-angled polygon" because people do not use "three-sided" to mean "three-angled."

§8.4 THE SENTENTIAL PRIMACY THESIS

The Gricean analysis does have its defects, however. For starters, (1) is incomplete as a theory of meaning because it tells us only whether or not an expression means *that* something is the case. Hence (1) cannot assign appropriate meanings to subsentential expressions, including *morphemes* ("non-," "hyper-," "-able"), individual *words* ("The," "cat," "is," "uncatlike"), *phrases* ("the cat," "on the mat"), and *clauses* ("that the birds sing," "when the sky clears"). Let us call subsentential expressions *subunits*, for short. The Gricean analysis correctly rules that subunits do not meant that p for any "p." For people do not use them to mean that anything is the case. For the same reason, though, (1) does not have the resources to tell us positively what "-able," "cat," and "when the sky clears" do mean.

14 The same problem affects Horwich's (1998a: 138–43) idea that the acceptance property for certain words is our regarding certain sentences containing these words as true ("implicit definitions"). The meaning of "true" may be connected with our regarding "The proposition that p is true if p" as true. But we will regard this sentence as true only if we regard the word "true" as having a certain meaning, the meaning we are trying to explain ("according with the facts" rather than "rightful," "accurately shaped," or "faithful"). If a sentence contains a meaningless word, we will not regard it as true (unless the word is mentioned). So it is hard to see how regarding sentences as true can give words meaning.

A radical Gricean might dig in and insist that subsentential words simply have no meaning.[15] It is implausible enough to claim that "blue" and "green" are like "elub" and "neerg" in being meaningless, but it is particularly hard to accept that "The sky is blue" is meaningful but not "the blue sky." Worse yet, the nihilist conclusion would undermine the fundamental principle of the compositionality of meaning. Because "gleft" is meaningless, any phrase, clause, or sentence in which "gleft" is used rather than mentioned is meaningless. So either the nihilist conclusion about subunit meaning would spread to sentences, or else the Gricean would face the difficult task of saying how "All boys are gleft" differs from "All boys are male."

Rather than rejecting (1) entirely, or finding illocutionary acts that subsentential units are used to perform, or denying the existence of subunit meaning, Griceans adopted the *primacy of sentence meaning*, the thesis that *the meaning of words is derived from the meaning of sentences.*[16] More specifically, Griceans believed that word meaning could be defined in terms of sentence meaning in such a way that what a particular subunit means is determined by what sentences containing it mean. Griceans added the proviso "where e is a sentence" to (1), and used it to define the meaning of sentences. They then claimed that subunits get any meaning that they have entirely from the fact that they are parts of meaningful sentences, and searched for definitions that would assign meanings to individual words solely on the basis of the meanings of the sentences in which they appear.

Many theses assigning priority to sentences are not concerned with the derivation of word meaning from sentence meaning. Frege (1884: §60) and Wittgenstein (1922: 3.3) maintained that "a word stands for something only in the context of a sentence." Wiggins (1971: 16) and Evans (1982: 102) similarly held that "the understanding of a word is manifested only in the understanding of sentences." Chierchia and McConnell-Ginet

15 This was suggested by Andrew Milne.

16 Cf. Grice 1969: 148–50; Schiffer 1972: 6, 166; 1987a: 92, 214–16; Bennett 1976: 16–22, 212–21, 272–6, 280–4, Loar 1976a: §1; 1981: §9.9; Blackburn 1984: 24, 129, 180; and Avramides 1989: 4–5. See also Bentham 1816: 188; 1843: 322; Frege 1884: §60; Ryle 1957: 248–9; Ziff 1960: 44–5, 141, 151, fn. 2, 160–1; Austin 1961; Alston 1963a; 408–10; 1964a: 36–9; 1971: 35–6; 1974: 33; 1977: 17; 1994: 48, 31; Kretzmann 1967: 390 (describing Destutt De Tracy); Armstrong 1971: 428; Davidson 1967: 451, 454; 1973: 127; 1977: 220; Danto 1975: 13–15; McDowell 1978: 308–9; Harrison 1980: 193–6; Hungerland & Vick 1981: 47, 58, 66, 135 (describing Hobbes); Chierchia & McConnell-Ginet 1990: 61–2, 152, 349; Vanderveken 1990: 7; Neale 1992: 555, fn. 68; Brandom 1994: 79–84; Hugly & Sayward 1995. Compare and contrast Dummett 1973: 3–4; 1976: 72, 76; and Peacocke 1986: 63–4. Contrast Grice 1968: 129–31.

(1990: 349) claimed that "word meaning must be able to provide an appropriate finite base for an adequate recursive theory of indefinitely many sentential meanings." And I cite evidence in §14.1 that ideas occur only as parts of thoughts, which implies that ideas expressed by words must occur as part of ideas expressible by sentences. These theses can all be accepted even if it is denied that word meaning can or must be defined in terms of sentence meaning.

The Fregean maxim that *words have meaning only in sentences*, which may be called the *sentential confinement thesis*, seems to me to be plainly false, at least when it concerns *Sinn* rather than *Bedeutung*. Those words forming the title to this chapter are meaningful, for example, even though they are not in sentences. The first word in the title means "basic," and the whole title means "basic word meaning." Meaningful words also occur outside of sentences in tables of contents, indexes, and catalogs; on labels and signs; and as instructions in musical compositions. But even if the sentential confinement thesis were true, it would not entail or support the thesis that *words derive their meaning from the meanings of sentences*. Conversely, sentential primacy would not entail sentential confinement. Hence the fact that meaningful words are not confined to sentences does not suffice to refute the thesis that word meaning can be defined in terms of sentence meaning.

Pro

Wallace (1977: 321–4) claimed that the primacy thesis "articulates an assumption which guides our thought on a range of philosophical problems, and guides it well." He had in mind Russell's theory of descriptions, and the reduction of number theory to set theory. Wallace assumed that it did not matter whether we identify the number three with one set or with another as long as the axioms and theorems of the resulting axiomatization are the same. Since Wallace took the sentential primacy thesis to predict this result, he used it to support the primacy thesis. But when we are concerned with semantics, it does matter whether we say that "3" means "$\{\{\{\emptyset\}\}\}$" or "$\{\{\emptyset\}\} \cup \{\{\{\emptyset\}\}\}$" (where $\emptyset$ is the empty set). Indeed, any theory saying that "3" *means* "$\{\{\{\emptyset\}\}\}$," or that there is no difference between meaning "$\{\{\{\emptyset\}\}\}$" and meaning "$\{\{\emptyset\}\} \cup \{\{\{\emptyset\}\}\}$," is false, nay preposterous. Russell's theory of descriptions attempted to explain the meaning of any definite description by giving a quantificational paraphrase of any sentence containing it. Sentences of the form "The F is G" are paraphrased as "There is a unique

F and it is G." The claim that such paraphrases have the same meaning or even truth conditions is controversial, however. More importantly, the issue at hand is whether "The F" derives its meaning from "The F is G," whatever that sentence might mean, or whether the latter derives its meaning from the former. Neither Russell nor Wallace addressed this question.

The Gricean *deduces* the sentential primacy thesis from theoretical premises. Grice's guiding principle is that word meaning is determined by cognitive speaker meaning. But cognitive speaker meaning is necessarily sentential. The only way that these two principles can be true is if the meaning of subunits can be defined in terms of the meaning of sentences. Such a transcendental deduction of the sentential primacy thesis provides little reason to believe that it is true. For if meaning that p is the only relevant kind of speaker meaning, as the Gricean maintains, then there cannot be any independent evidence for the guiding principle that *all* word meaning – subsentential as well as sentential – is determined by speaker meaning. The part of that premise concerned with subsentential meaning needs just as much support as the sentential primacy thesis that is inferred from it. Grice's reasons for adopting the sentential primacy thesis are more strategic than evidential. The deduction tells him he has to *hope* that the sentential primacy thesis is true, otherwise he will have to abandon his theoretical program. Parallel observations apply to truth conditionalists[17] and inferentialists who adopt sentential primacy for similarly transcendental reasons.[18]

Is there any evidence supporting the thesis that the meaning of terms derives from the meaning of sentences? Alston (1974: 44–5; 1977: 18) cited the fact that "sentence meanings constitute the major source of data about semantics." It would be difficult to find a rigorous and non-question-begging justification for such a methodological claim, especially since subunits include phrases and clauses. For a linguistic theory it is no more important to correctly predict the meaning of a novel sentence like "The president made chartreuse Jell-O" than it is to correctly predict the meaning of a phrase like "The president's chartreuse Jell-O." And only verificationism would lead from a premise about possible evidence for semantic theories to a conclusion about the nature of meaning.

17 See, e.g., Peirce 1931–5: §§5.476, 5.411–34; Quine 1953: 42; Davidson 1967: 451; Kretzmann 1967: 395; Lewis 1969: 198; 1975: 20; Wiggins 1971: 17–18; Dummett 1973: Chapter 1; Wallace 1977; and Evans 1982: 102; 1985: 72–3. Contrast Fodor & Lepore 1992: 40–8; 216–18.

18 Block 1993: 1; Boghossian 1993; Lance & O'Leary-Hawthorne 1997: 8–10, 370.

Others cite putative facts about how we learn the meanings of words. Thus Bennett (1976: 19) supports sentential primacy by observing that "to learn the meaning of any word one must learn its use in some sentences." Blackburn (1984: 24) similarly proffers that "if we are working our way into a language from scratch, in principle there will be no sharp division between learning to isolate words, and learning the meanings of statements." But such premises about epistemological priority imply nothing about definitional priority, and are perfectly compatible with the opposite claim that sentential meaning must be definable in terms of subsentential meaning. Schiffer (1972: 6) cites a syntactic truism: "[S]omething is a part-utterance type only if it is the sort of thing which when combined in certain ways with certain other things yields a whole-utterance type." Nothing follows from this about meaning.

Alston also offers the premise that "a sentence is the smallest linguistic unit that can be used to perform a complete action that is distinctively linguistic" (1964a: 33), "the minimal vehicle for a self-contained speech act" (1974: 45). Dummett (1976: 72) uses nearly identical language. The paradigm of "complete linguistics acts" for Alston and Dummett are illocutionary acts like asserting. Assertion does indeed require a sentence. But acts like expressing the concept "dog," referring to dogs, and ascribing the property of being a dog, all of which people can use the word "dog" to perform, are as distinctively linguistic as assertion and seem perfectly "complete" in their own right (cf. Hugly and Sayward 1995: 410–13). Such acts would be incomplete if you were trying to state that dogs bark, just as the act of playing a middle C would be incomplete if you were attempting to play an entire scale. But expressing the idea of a dog is no more incomplete in and of itself than the playing of middle C. One might try arguing that assertion is more complete than the expression of concepts because one cannot express the concept of a dog without asserting something about dogs. Even if this premise were true (the use of words in titles and labels suggests otherwise), it would establish no priority or completeness for sentence meaning. For it is no less true that one cannot assert anything without expressing concepts, and thereby referring to something and describing it. A use theory of meaning like Alston's leads to sentential primacy only if the class of acts that words can be used to perform is arbitrarily restricted.

Dummett claims that "the sense of a word or of any expression not a sentence can be understood only as consisting in the contribution which it makes to determining the sense of any sentence in which it may occur"

(1973: 4). It may be granted as obvious that a word does contribute its meaning to any sentence in which it occurs, and that any expression that makes no contribution to the meaning of any sentence is meaningless. But what is Dummett's evidence that the meaning of a word *consists in* its contribution to sentence meaning? As Hugly and Sayward (1995: §4) observe, it is equally true that a sentence contributes its meaning to any singular term in which it appears. "Oswald killed Kennedy" makes a definite contribution to the meaning of "the place where Oswald killed Kennedy," "the fact that Oswald killed Kennedy," and so on. It does not follow that the meaning of a sentence consists in its contribution to the meaning of singular terms.

One of the oldest and most persuasive arguments for the sentential primacy thesis was spelled out by Armstrong.

> We must now take notice of Locke's major error. Locke . . . develops his theory in terms of words, not sentences. Words are said to be signs of "ideas" in the mind of the utterer of the word. But if we want to develop a theory of meaning in terms of communication, must we not concentrate on the meanings of *sentences* rather than words? If somebody utters a mere part of a sentence, as a word usually is, what does that communicate? In general, very little. Suppose I just say, "Horse." Has anything much been communicated? In general, a complete sentence is needed to get something across. The *unit of communication* is the sentence. And so a theory of meaning that is based on the communicative function of language must treat the sentence as in some way semantically fundamental. An account of word meaning will have to be given in terms of the contribution of words to the meaning of sentences. This paper will not get very far beyond the meaning of sentences. (Armstrong 1971: 428)[19]

On my account, word meaning in living languages is dependent on the communicative function of language. As I shall put it in Chapter 9, communication is the common interest that sustains the speaker meaning conventions that constitute word meaning. So the bare consistency of my theory should show that Armstrong's argument is a non sequitur. To base a theory of word meaning on the communicative function of language is not necessarily to identify what an expression means with what it communicates. Nor does a communication-based theory require defining word meaning in terms of sentence meaning. To combine a Lockean

19 Cf. Quine 1967: 306; Davidson 1973: 127; Alston 1974: 45, fn. 21, 46; 1980: 124; Brandom 1994: 83; Hugly & Sayward 1995: 417. Hungerland & Vick (1981: 47, fn. 20) attribute the argument to Hobbes. Contrast Ockham, *Summa Logicae I*; Loux 1974: 2.

view of word meaning with the view that word meaning is based on communication, it suffices to observe that (1) communication involves the communication of thoughts (Chapter 5); (2) thoughts are complex entities with parts (Chapter 14); (3) the words in a sentence are used to express the parts of the thought that the sentence is used to express and communicate (Chapter 10); and (4) what an expression means is determined by what it expresses (Chapters 7 and 21). This accounts for the contributions of words to the meanings of sentences, but does not imply that sentence meaning is in any way fundamental. Nor does it imply that individual subunit meaning is fundamental.

I believe that Armstrong's argument is a nonstarter as well as a non sequitur. True, individual words are not generally used to communicate information, beliefs, desires, or even thoughts. Hence they may not communicate "much." But individual words are used to communicate *ideas*. By uttering "horse," speakers of English do something by which they express the idea of a horse and from which hearers recognize that they are expressing that idea. Hence, by Definition 5.3, people customarily use the word "horse" to communicate the idea of a horse. In general, we can communicate anything that we can express. The communication of ideas is generally important to us only insofar as it is part of the process of communicating thoughts and information. But that does not mean that individual words do not communicate anything. And it is not true that the only reason we use words and express ideas is to say things, as titles and labels again show.

Hugly and Sayward (1995: 418) postulate that any system of expressions that did not contain sentences (that is, expressions used to say that something is the case) would not count as a language. This is a reasonable conjecture, but I am inclined to think that it is false. A symbol system consisting of nothing but names of, or singular terms for, propositions would be quite serviceable. To assert that the sky is blue, we could utter the name of that proposition while nodding our head. Even if languages must contain sentences, it still does not follow that the meanings of words derive from the meanings of sentences. Indeed, one could just as well strengthen their postulate to say that a system of expressions does not count as a language unless it contains sentences whose meaning is derived from the meanings of their components. Their deeper premises that languages are means of communication and that communication is a matter of saying something are perfectly compatible with this stronger postulate, which reverses the order of derivation.

We have seen that the reasons offered in support of the sentential primacy thesis provide little evidence of its truth. The arguments against the theory are more substantial. First, *the primacy thesis postulates diversity where we should expect uniformity.* The views of Grice and his followers seem to imply that the meanings of sentences and the meanings of phrases are very differently related to speaker meaning. Their views imply, for example, that although the meaning of "The ball is red" is determined by what people mean by that expression, the meaning of "the red ball" is not determined by what people mean by that expression. It would be hard to explain such a difference. The fact that one is a sentence and the other a phrase seems irrelevant.

Second, it is a platitude that *the meaning of a sentence is generally determined by the meanings of the words of which it is composed*. By getting the order of determination backward, the primacy thesis conflicts with compositionality. It would be difficult to dismiss compositionality as an illusion. For it seems essential that a finite theory account for the infinity of meaningful expressions in any natural language, and that such a language be finitely learnable.[20]

Dummett (1973: 4) tried to resolve the apparent conflict between compositionality and sentential primacy by distinguishing between the order of explanation and the order of recognition. While the meaning of a sentence is recognized on the basis of the meanings of the words it contains, he said, the meaning of a word is explained on the basis of the meanings of the sentences in which it appears. But it is precisely the latter claim that seems backward: the sentence "They saw wood" has the meanings it does *because* the words "they," "saw," and "wood" have the meanings they do, and not vice versa. The same goes for any sentence in which they appear. The order of recognition, by contrast, can go either way, as it can with most causal or explanatory connections: sometimes we figure out what a word means on the basis of what a sentence containing it means; sometimes we go from word meaning to sentence meaning.

20 Cf. Quine 1967: 306; Wiggins 1971: 24–5; Schiffer 1972: 6; Alston 1974: 43–4; 1977:18; Bennett 1976: 19; Avramides 1989: 4, fn. 18; Fodor & Lepore 1992: 54–5, 210, n. 5; Hugly & Sayward 1995: 411, 415. Grice himself asserts that sentence meaning is determined by word meaning in Grice 1968: 235–6, and in the version of "Utterer's Meaning and Intention" published in 1989 (see Grice 1989: 87–8), but not in the version published in 1969. Neale (1992: 555, fn. 68) notes that Grice accepted compositionality, without explaining how Grice's definition of word meaning could allow it.

Alston correctly observes that as long as the orders of derivation are different, the claim that one thing is derived from a second does not conflict with the claim that the second is derived from the first.

> *The fact that a given sentence has a certain meaning* is derivative from *facts about the meanings of its constituent words*; whereas the *general concept of word meaning* is derivative from the *general concept of sentence meaning*. It is in the order of derivation (or explanation) of particular semantic facts that word meaning is prior, while it is in the order of derivation (or analysis) of semantic concepts that sentence meaning is prior. (Alston 1980: 133, n.2)[21]

However, analyzing the general concept of meaning is, for us, just the process of saying what it is for w to mean μ, for any particular word w and meaning μ. If the meaning of any sentence containing w derives from the meaning of "w," then w cannot derive its meaning from the meanings of sentences containing it.

If one maintained, as I do, that words have something to contribute to the meaning of sentences that is independent of the meaning of sentences, then there would be no difficulty in holding both that sentence meaning is compositional and that the meaning of a word is what it contributes to the meaning of a sentence. The difficulty arises because the Gricean (along with the truth-functionalist and inferentialist) is forced to hold that *words get their meanings entirely from the fact that they are parts of meaningful sentences*. It is this implication that cannot be maintained if sentence meaning depends causally on word meaning. According to the sentential primacy thesis, the property of words that gives them meaning is the fact that they are parts of meaningful sentences. But the mere fact that they occur in meaningful sentences cannot be what words contribute to the meanings of sentences. There must be some independent conventions governing words that give them meaning.

If advocates of the primacy thesis had shown how to define subunit meaning in terms of sentence meaning, there would be less objection. But to my knowledge, *no one has succeeded in showing how meanings could be assigned to individual words on the basis of the meanings of the sentences in which they appear.* Only a few have even tried. It is not enough to say that the meaning of a subunit is what it contributes to the meaning of a sentence, unless one can specify something that a subunit does so contribute and describe it as something other than "its meaning." It is also not enough to say that the meaning of a term is a function of its use in a sentence. That

21 See also Alston 1974: 43–4; 1994: 31.

might be true only because w means μ iff w is used to mean μ in some sentences, which would result in a circular definition of word meaning.

Block (1993: 6) attempts to substantiate the primacy thesis within an inferential role semantics. He identifies the meaning or "semantic value" of a sentence with the ordered pair whose first member is the set of inferences to that sentence, and whose second member is the set of inferences from that sentence. He then suggests that the meaning of a word be identified with the set of semantic values of sentences containing that word. We can add that w means μ iff w has the same meaning as μ, provided that we know the meaning of μ. Given these definitions, there is (trivially) a function from the semantic values of a sentence's components to the semantic value of the sentence. Block concludes that the inferential semanticist need not give up compositionality. But the function that yields the semantic value of a sentence from the semantic values of its components is not such that the meaning of a sentence contains in any sense the meanings of its components. The meaning of a word, on Block's proposal, is not something that it contributes to the meaning of a sentence. On the contrary, the meaning of a sentence is contained in the meanings of its component words. Furthermore, Block did not explain how the meaning of a phrase depends on the meanings of its components. There appear to be significant relations between the semantic values that Block's theory assigns to "red ball" and "red," for example. But the theory provides no explanation of why this is so.

Block's "compositional" function does not even enable us to do the basics of formal semantics, such as calculating the truth value of a sentence on the basis of the extensions of its components. Indeed, it is not clear what property of an ordered pair of inference sets would make a sentence possessing it true, or what property of a set of ordered pairs of inference sets would make a word denote an object or set of objects. To fully grasp the latter problem, note that if a word appears in a sentence, it also appears in the negation of that sentence, and in some self-contradictions. So the semantic value of "cat" on Block's theory contains the inferential role of "cats are animals" as well as the inferential role of "cats are not animals" and "cats are not cats," whatever the latter might be. There are technical problems, moreover, with Block's formal apparatus. For example, the theory makes it impossible for any language to add a word that has the same meaning as an existing word. Adding a word to the language will expand the set of sentences that existing words appear in, thus changing the semantic value of every word in the language. Finally, Block's theory does not begin to tell us what it is for a sentence or a word to have a

meaning, because it tells us nothing about what determines the inferential roles of sentences. Block has not provided a satisfactory theory of either word meaning or sentence meaning.

Alston (1964a: 37) suggested that a subunit w means μ provided w can be replaced by μ in any sentence without changing the meaning of that sentence. This supplement attempts to define word meaning indirectly in terms of illocutionary acts, assuming that sentence meaning has already been directly so defined. But as Alston realized, the fact that words can be mentioned as well as used scotches this idea: "kitty" cannot be replaced by "kitten" in "'Kitty' is a five-letter word" or in "The word *kitty* is ambiguous" without changing the meaning of those sentences. The presence of the same expression in different grammatical categories is also problematic: "fall" means "drop," but it cannot be replaced by "drop" in "Fall comes after summer." We can view the ambiguity of "fall" as the limiting case of amphiboly.

The mention and amphiboly problems can be avoided by weakening "all" to "some," as in Alston's definition of sentence meaning.

(4) w means μ iff w can be replaced by μ in some sentences without changing the meaning of those sentences.

As written, (4) is too weak. It implies that "ph–" has a meaning, namely "fffff," since the latter letter sequence can replace the former in "The balloon went 'ph–'" without changing the meaning of the sentence. And it follows that "mise" means "miss," since the latter can replace the former in "The first premise is false" without changing the meaning of the sentence. These problems can perhaps be avoided by requiring that w be a grammatically defined unit of the language, such as a word or word segment.

But (4) is also too strong. One problem is the existence of meaningful terms in other languages. The German word "hut" means "hat," but there is no sentence in which these two words can replace each other without changing the meaning of the sentence. Even when we stay within one language, (4) makes it impossible for two terms to be synonyms if they fail to occur in the same context. It would be easy to construct a language or code in which synonyms did not or could not replace each other. Imagine a written language that is like English, except for having the strange but strict orthographic rule that capitalized words cannot occur with uncapitalized words. Then "cat" would be synonymous with "CAT" despite the fact that one cannot replace the other in any sentence. Or imagine a language like English except that "every cat" must be followed by "is," and "All cats" by "are." In natural languages, though, it is hard

to find synonyms that never replace each other. But that hardly seems to be an essential feature of natural languages, and I would be amazed if it were universal. Indeed, my searches have turned up no case in which the suffixes "-able" and "-ible" can replace each other in English, despite their having the same meaning. A separate problem is how the Alstonian proposal can yield the nearly tautological result that "cat" means "cat." If we secure this by allowing that an expression may replace itself, we also get the unwanted result that "Gorf" means "Gorf," since that expression can replace itself in "'Gorf' is meaningless." A related problem is what account the Alstonian can provide for "w means something" without ruling either that a meaningful word must have at least one synonym, or that "Gorf" is meaningful.

Alston (1971: 40) later realized that word substitution without a change in sentence meaning could be only a *test* of sameness or difference of word meaning, not a *definition* of word meaning. For the difference in meaning between "John blew out a tire" and "John blew out a candle" only establishes a difference in meaning between "tire" and "candle" if we assume that "John blew out a" has the same meaning in both sentences. Since the latter expression is itself a subsentential unit, the assumption concerns word rather than sentence meaning. Similarly, the synonymy of "Vixens purr" and "Female foxes purr" only establishes that "vixens" and "female foxes" mean the same thing if we assume that "purr" does not undergo a compensating change in meaning between the two sentences. So any formula providing necessary and sufficient conditions for word meaning in terms of substitution would have to employ the notion of word meaning being defined.

Grice's (1968: 130–1) attempt to assign meanings to individual words on the basis of the sort of meaning possessed by sentences – namely, meaning that p – was sketchy, incomplete, and complicated. We need not go into the forbidding details, because Grice's analysis had a fundamental defect that can be grasped by examining the following simple schema.

(5) "e" means "P" iff the sentence "Something is e" sometimes means that something is P.

Since the sentence "Something is a vixen" means that something is a female fox on some occasions, and on others means that something is a shrewish woman, (5) correctly rules that "vixen" is ambiguous, meaning "female fox" and "shrewish woman." But (5) only works because meanings have already been assigned to the words "something" and "is." We might try assigning a meaning to "something" using (6).

(6) "e" means "something" iff the sentence "e is a vixen" means that something is a vixen.

But (6) only works if we have already assigned a meaning to "vixen." If we start from scratch, we get nowhere: the fact that the sentence "bliff blog e" means that something is a female fox gives us no reason to infer that it is "e" that means "female fox" rather than "blog." The problem with Grice's procedure is that without having assigned meanings to some words already, it cannot assign meanings to any words on the basis of what sentences containing those words mean. So Grice's procedure provides no support for the sentential primacy thesis.[22]

Bennett's (1976: 216–41) discussion of how we might determine the meaning of subsentential units on the basis of the known meanings of a finite number of sentences is not intended to provide a definition of word meaning, but it suggests as plausible a definition as I have seen for names and predicates:

(7) w means "n" iff all sentences containing w mean something about n's.[23]

This is difficult to apply when w is ambiguous, as are nearly all terms in a natural language. For example, is the sentence "Bill shot a vixen" about female foxes, shrewish women, or both? It seems incorrect to say "both," but we have to if the Bennett formula is to yield the correct result that "vixen" means both "female fox" and "shrewish woman" in English. It would seem necessary to say that sentences containing "vixen" *in one sense* mean something about female foxes, and *in another sense* mean something about shrewish women. But this would make the Bennett definition circular.

For different reasons, (7) is too weak as a definition. Waiving ambiguity, all sentences containing "female foxes" mean something about foxes; but "female foxes" does not mean "foxes." Similarly, all sentences containing "he," "Mary's husband," or "The man named George Bush" mean something about males, but these terms do not mean "male." The

22 As Grice (1968: 135) noted, his formulation also suffered from Frege's problem. See note 23.

23 Cf. Chisholm 1958: 239; Ziff 1960: 62–5, 89–90; Alston 1963a; 1964a: 36–9; Grice 1968: 132–6; Loar 1981: 225–6; Blackburn 1984: 129; and Chierchia & McConnell-Ginet 1990: 87. Contrast Loar 1981: 234. If the "about" in (7) is treated relationally rather than intentionally (§6.3), the Bennett procedure would have an additional headache in the form of Frege's problem (§23.1): since sentences "about" water drops will also be sentences "about" H_2O drops, the relational variant of (7) would entail that "water" and "H_2O" are synonyms. Grice's (1968: 132–6) formulation generated Frege's problem for the same reason.

existence of quotation makes Bennett's formula too strong: "vixen" means "female fox" even though the sentence "'Vixen' is a five letter word" does not mean anything about female foxes. So does the existence of terms like "fox terrier": "fox" means "fox" even though sentences like "I own a fox terrier" do not mean anything about foxes. Changing "all" to "some" makes the suggested definition much too weak: "vixen" does not mean "cat" even though the sentence "A cat is not a vixen" means something about cats.

A natural reaction here is to think that with suitable patience and intelligence, the Bennett formula could be clarified and modified to avoid such problems. But I think there is a way to demonstrate that the primacy thesis is false. The basic premise of the refutation is this: *the assignment of meanings to sentences does not determine a unique assignment of meanings to words*. The refutation is similar to Quinean arguments for the "inscrutability of reference,"[24] except for focusing on meaning rather than reference. Consider a language containing just three elementary expressions "a," "b," and "L," and the two sentences "aLb" and "bLa." The primacy thesis implies that information about the meanings of sentences suffices to determine the meanings of terms in the language. This fails in the case at hand. There are too many subunits given the number of sentences in the language. In effect, the number of simultaneous equations is insufficient to solve for the number of variables present. Suppose we are given the following:

"aLb" means that Dick loves Jane.
"bLa" means that Jane loves Dick.

The two meaning statements do not suffice to fix the meanings of the three terms. The given information is compatible with the following two hypotheses: (H_1) "a" means "Dick," "b" means "Jane," and "xLy" means "x loves y"; (H_2) "a" means "Jane," "b" means "Dick," and "xLy" means "y loves x."

This symbol system is quite simple, of course. Even if it is too simple to count as a language, it still casts doubt on the primacy thesis: what reason is there to think that true languages differ in the relevant respect? This is especially pressing since the above symbol system can be complicated considerably without there being enough sentences relative to the number of terms for sentence meaning to determine term meaning.

24 See Quine 1969a; Lewis 1975: 18–21; Wallace 1977; Davidson 1977; and Nelson 1992: 134–5. Cf. Loar 1981: §9.9 and Schiffer 1987a: 254.

We can add quantifiers, variables, and sentential connectives. We can add as many "names" ("a," "b," "c," "d," etc.) and "two-place predicates" ("L," "M," "N," "O," etc.) as we like, as long as a predicate is prohibited from occurring between two occurrences of the same name or variable.[25] We can even add one-place predicates ("A," "B," "C," etc.) as long as they can occur only with variables and a special category of names ("α," "β," etc.). Finally, we can add a finite set of meaningless units (letters or speech sounds) out of which each name and predicate is formed. The expressive capacity of this symbol system will scarcely be restricted by its grammatical idiosyncrasies, since its predicates may include some meaning "instantiate" or "is a member of," and its names may include some meaning "the property of loving himself" or "the set of self-identical objects." Moreover, the names in the two syntactic categories need not differ in meaning. The resulting symbol system differs markedly from English and other natural languages, to be sure. But I see no reason to deny that the language contains meaningful sentences and terms, which is all we need to demonstrate that subsentential meaning cannot be defined in terms of sentence meaning.

The simple symbol system consisting of "aLb" and "bLa" illustrates another way in which assignment of meanings to sentences does not determine an assignment of meanings to words. A third hypothesis (H_3) compatible with the two meaning assignments just described is that "a," "L," and "b" are meaningless. Just as we cannot infer from the premise that "red" and "black" have meanings to the conclusion that "r," "e," "d," and so forth have meanings, so we cannot infer from the mere fact that "aLb" and "bLa" have meanings to the conclusion that "a," "L," and "b" do. Even if the arrangement of the components of an expression enabled us to infer the meaning of the whole expression, that would not entail that the components have meanings, as "red" and "black" again illustrate. So the meaning of a word must consist in something more than the fact that its presence helps us to predict the meanings of sentences containing it. The presence of "r" in a certain position helps us to predict that a word means "red." But we do not and cannot infer that "r" *contributes something to the meaning* of the word. So even though we can infer the meaning of "aLb" and "bLa" from the arrangement of "a," "L," and "b," that does not entitle us to conclude that "a," "L," and "b" contribute anything to the meanings of "aLb" and "bLa." These expressions might be mere statement

25 If we add "'aLa' means that Dick loves Dick," we can infer hypothesis H_1. Three linear equations can be solved for three variables.

constants, as in propositional logic, despite consisting of three letters rather than one. Given our training in quantification theory, it is hard for us to think of "aLb" as anything other than a complex expression consisting of two individual constants and a predicate constant. But suppose that in place of "aLb" and "bLa" we have "ɨ" and "ɨ," which also have three distinguishable components (dot, diagonal stroke, vertical stroke). The hypothesis that the components have separate meanings would not be so intuitively compelling.

Suppose we were translating a foreign language, and somehow determined that a wide range of sentences containing w meant something about n's, and that these sentences had nothing else in common besides containing w. Then surely it would be a reasonable hypothesis to conclude, tentatively, that w means "n." This hypothesis would be confirmed if we found natives pointing to n's and uttering w. If we then found other sentences containing w that were not about n's, we might hypothesize that w is ambiguous, or we might look for evidence of something like quotation. I do not deny the reasonableness of this method of inquiry. The fact that it is reasonable does not, however, entail the sentential primacy thesis. This is particularly clear since the method of inquiry in question is inductive, producing tentative hypotheses subject to future refutation. The proposition that we verify claims about word meaning on the basis of evidence about sentence meaning does not, *pace* verificationism, entail that word meaning is definable in terms of sentence meaning.

Finally, we will observe in Chapter 10 that the sentential primacy thesis would make it difficult, if not impossible, for a use theory to account for the infinity of unused but meaningful sentences that any natural language contains.

§8.5 THE BASIC NEO-GRICEAN ANALYSIS

Grice suggested that "e means that p" could be defined by "people use e to mean that p." Since only sentences mean *that* anything is the case, this analysis does not provide a complete definition of word meaning. The failure of the sentential primacy thesis shows that the Gricean analysis is unlikely to be extended to provide a satisfactory analysis of meaning in general.

Even where it applies, to sentences, the Gricean analysis is defective. People often use "That's great!" to mean that something is the very opposite of great, and use "It weighs a ton" to mean only that something

is heavy. Yet "That's great!" does not mean that something is the very opposite of great, and "It weighs a ton" means more than that it is heavy. The Gricean analysis thus wrongly turns figurative use into literal meaning.[26] Moreover, an expression e might be used only in science fiction stories. Then people never use e to mean *that* anything is the case, in which event e means nothing according to the Gricean analysis. So the Gricean analysis wrongly turns exclusively fictional use into nonsense.[27] The same goes for nondeclarative sentences[28] and dependent clauses. Finally, since "He sings" means "He sings," the Gricean analysis entails that speakers conventionally use "He sings" to mean that he sings. That sounds fine. Unfortunately, *meaning that he sings* denotes different actions on different occasions. When S is referring to Pavarotti, he means that *Pavarotti* sings. When S is referring to Glenn Gould, he means that *Glenn Gould* sings. There is no one belief that people commonly use "He sings" to express (or to produce in others). The converse defect is even clearer. People often use "He sings" to mean that Pavarotti sings. But "He sings" does not mean "Pavarotti sings."[29]

The four problems with the Gricean analysis – figurative use, fictional use, pronouns, and subsentential meaning – can be solved at one stroke by shifting to *cogitative* speaker meaning. As explained in Chapter 2, people use "That's great!" to mean cogitatively "That is great!" while meaning cognitively that it is the very opposite of great. I presume that no one has ever used "Man was establishing his first base on the Moon" to express a *belief* that man was establishing his first base on the Moon. But people have used the sentence to mean that (as Arthur C. Clarke did in *2001: A Space Odyssey*). *Meaning "He sings"* is a single action that people commonly use "He sings" to perform. By performing this action, people sometimes express the belief that Pavarotti sings, sometimes the belief that Gould sings, and so on ad infinitum. Finally, people use "vixen" to mean something, viz., "female fox," even though they do not use "vixen" to mean

26 The same problem is presented by Schiffer's (1972: 119–20) example of people who use "grr" to mean that they are angry, and Carr's (1978a) example of "Can you pass the salt?" used to mean that you are to pass the salt. See also Loar 1976b; 1981: 256–7, and Posner 1980: 169–80. I discuss figurative use, implicatures, and indirect speech-acts at length in Davis 1998.

27 Blackburn's (1984: 125) suggestion that for e to mean p is for people to use e to *display that p*, i.e., to perform an action habitually taken by a group as a natural sign that p, also suffers from this defect. Contrast Meinong 1910: 29–33.

28 Cf. Landau 1984: 215–16.

29 Cf. Cartwright 1962: 92–5; Chierchia & McConnell-Ginet 1990: 153–4; Neale 1992: 553–4; and S. Davis 1994.

that anything is the case. When cogitative meaning replaces cognitive, the Gricean analysis becomes neo-Gricean:

(8) Expression e means μ iff people use e to mean μ.

This solution was closed to Grice and his followers. Schiffer and Grice defined cogitative speaker meaning in terms of word meaning (see §4.5), and so could not have defined the latter in terms of the former. Bennett (1976: 24) held that the distinction between cogitative and cognitive speaker meaning was a negligible "nuance." Note that (8) does not rely on the primacy of sentence-part meaning, any more than it relies on the primacy of sentence meaning. It says that the meanings of words in general, sentences as well as their parts, are determined by what people use them to mean.

Not all use theorists who followed Wittgenstein and Alston adopted the sentential primacy thesis. Carruthers (1989: Chapter 10) and Woodfield (1997), for example, focus on general terms and the act of *classification*. Their work suggests the following alternative to (8):

(9) Expression e means "G" iff people use e to classify things as G.

Thus "vixen" means "female fox" because people use "vixen" to classify things as female foxes. We do this when we say, for example, "That is a vixen." This version of the use theory can be extended naturally to singular terms by focusing on the act of *reference*. The corollary would say that e means "S" iff people use e to refer to a certain object as S. Thus "WWII" means "World War II" because people use "WWII" to refer to a certain war as World War II. Even though they are markedly different from asserting, warning, and promising, classification and reference are illocutionary speech acts. It is plausible that every act of assertion consists of at least one act of reference combined with an act of classification.

While (9) and its corollary are quite plausible when they apply, (8) has the advantage of *generality* and *fundamentality*. The greater generality of (8) becomes apparent when we try to extend the classification/reference theory beyond singular and general terms. Sentences can be accommodated by using the speech act of *saying* (see (3) in §8.2). But what parallel speech acts are verbs, adverbs, logical constants, prepositions, infinitive phrases, and clauses used to perform? It makes no sense to say that people use "or" to refer to or classify things as "or," for example, even though they use "or" to mean "or." The greater fundamentality of (8) comes out when we try to explain what it is to use e to classify things as G. Surely a minimum requirement is that we use e to express the thought that something is G.

A speaker classifies Φ as G by saying "Φ is e" only if S means "G" by "e." Using e to classify Φ as G involves meaning something by e and expressing a particular thought about Φ.

Despite its considerable advantages over the other analyses we have examined, formulation (8) is still much too weak. People often confuse "entomologist" and "etymologist." As a result, people often mistakenly use "etymologist" to mean "entomologist." It does not follow that "etymologist" means "entomologist."[30] Similarly, if a group used "scarlet" in a code to mean a battleship, their usage would not affect the meaning of that English word at all. At the level of sentences, it is quite possible for an English sentence like "The fox chased the whale into the bullpen" to be used only by actors testing their voices, or by spies using codes, so that it is never used to mean what it actually means in English. In general, what a word means is not influenced by what particular speakers happen to mean by it, a phenomenon sometimes called the *autonomy* of word meaning.[31]

This problem can be avoided by insisting that word meaning is *conventional* speaker meaning, where a convention is an *arbitrary social practice* or *custom* (see Chapter 9).[32] Codes are deliberate departures from convention. And while many people do use "etymologist" to mean "entomologist," their usage is unconventional. It is not "the practice." If a linguistic mistake became conventional, it would no longer count as a mistake: the language would have changed. The use of sentences to test one's voice, or by spies in a code, is clearly unconventional. I will refer to the thesis that word meaning is conventional cogitative speaker meaning as the *basic neo-Gricean analysis*.

(10) e means μ iff it is conventional for people to use e to mean μ.[33]

30 Cf. Ziff 1960: 22–4; Cartwright 1962: 95; Chomsky 1965: 3; Stampe 1968: 171; Follett 1970: 4–6, 24; Alston 1974: 19; Tsohatzidis 1994: 2; Love 1994: 779. Contrast Newman (1962: 437: "A dictionary is . . . authoritative only to the extent that it accurately records what has actually been written and said") and Hall (1962: 434: "The function of grammars and dictionaries is to tell . . . what people actually do when they talk and write"), who, I suspect, were simply overstating their case.

31 Cf. Locke 1690: §3.2.8; Leibniz 1709: §§3.1–3.2; Ziff 1960: 22–4; Follett 1970: 4–6; Harrison 1980: Chapter 1; Davidson 1983: 274–9; and Avramides 1989: 74. Davidson also asserts that the meaning of an utterance does not determine its force or extralinguistic purpose, which is another type of autonomy; compare and contrast M. Green 1997.

32 See Grice 1989: 298; Neale 1993: 553; Avramides 1997: 80. In Grice 1968: 127, Grice revised his original tentative analysis by adding a proviso that foreshadowed the self-perpetuation condition of the definition of convention (Chapter 9 of this volume).

33 Compare and contrast Leonard 1929: 21–6, Chapter 9; Schiffer 1972: 154; 1982: 123; 1987a: 12; Loux 1974: 4; Fodor 1975: 106; 1998a: 9; Bennett 1976: 10, 16–17, 213;

When "μ_i" stands for words expressing i, this can be reformulated as saying that *e mean* μ_i *iff it is conventional for people to use e to directly express idea or other mental state i* (see §7.6). Since (10) uses the two key ideas stressed by a long tradition stretching from Aristotle through Boethius and Ockham to Hobbes and Locke (see §1.1), it could equally well be called the *neo-Aristotelian* or *neo-Lockean* analysis.

The neo-Gricean analysis applies to written words as well as to spoken ones. The rules for pairing letter sequences with speech sounds in English are satisfied equally well by "tow" and "toe." It is only because speakers conventionally use "tow" to mean "pull" rather than "foot digit," and "toe" to mean "foot digit" rather than "pull," that the written words have the particular meanings that they do and are unambiguous, unlike the spoken word /tō/, which is conventionally used to mean either "pull" or "foot digit." The analysis correctly rules that misspelled words do not mean what the speaker may have meant by them. The rules for pairing letter sequences with speech sounds in English would have been satisfied equally well if "although" had been spelled "altho" or "althow." The last two letter sequences are not words in English, however, because their usage has not become conventional. They are misspellings of "although." If a writer uses "althow," we will presume that he or she means "although." But the expression "althow" has no meaning in English. The neo-Gricean analysis assigns no meaning to "althow," because it is not conventionally used to mean anything. Individual speakers may use it to mean "although," but their usage is not conventional. When a writing system is first introduced into a language-using community, we can expect a period of variability in spelling. During this period, the written expressions will not be part of the language, and it will not be possible to say what the written expressions mean as opposed to what speakers mean by them. Only when the spellings become conventional will the written words become part of the language. When that happens, their meaning will be given by the neo-Gricean analysis. Written words get their meaning directly from conventional usage, not just from their association with spoken words (cf. §7.5).

Unlike the Gricean analysis, the neo-Gricean analysis correctly allows that *all* speakers of a language could misuse a word on occasion. For the

Cummins 1979: 352; Harrison 1980: 183–9; H. H. Clark 1983: 316; Blackburn 1984: 90, 112–13; Emmett 1988: 79–80; and Tsohatzidis 1994: 2–3. As we have seen, Alston (1964a: 43–4; 1974: 17–20; 1977:29; 1994: 45) championed the view that words have meanings because they are governed by rules, while specifying a different set of rules. Contrast Sellars 1969: 115.

fact that all people do something at a particular time does not entail that it is conventional to do so. A convention is more than a momentary, accidental regularity in action. The neo-Gricean analysis also enables us to explain why, despite the autonomy of word meaning, speaker meaning is the more fundamental phenomenon. It often happens, for example, that an individual uses an expression in a new sense, or uses a completely new expression. It is only when this usage "catches on," and becomes conventional, that the word acquires a (new) sense.

The neo-Gricean analysis is supported by the standard lexicographic practice of basing dictionary entries on citations. "The existence of an ongoing citation file is what distinguishes reputable, general dictionaries from purely derivative works," which must rely on other dictionaries or on personal linguistic intuitions (Landau 1984: 153). Representing the top of the scale, the editors of the *Oxford English Dictionary* spent twenty years gathering millions of citations. Dictionary editors since have generally focused on new and obsolescing words. Cowie, Mackin and McCaig (1983) gathered over 30,000 citations from the postwar period for their dictionary of current idioms. A citation records a sentence in which the target word is actually used, along with the source, date, and other information. The editors then determine what the author meant on that occasion. If the Gricean analysis were correct, a number of citations would suffice to establish what a word means. But in fact, the criteria are more strict.

> The four basic considerations in evaluating whether the citations justify the inclusion of a word are their number; the period of time covered; their geographic distribution; and the diversity of sources. One can compromise on one of the four criteria if the term in question meets the other three. For example, *streaking* lacked a sufficient period of time but was widely distributed, occurred in diverse sources, and existed in profuse numbers. (Landau 1984: 162)

> A single citation proves very little, but even two citations from different kinds of publications may prove a great deal. They prove that it is not a "nonce word" – a term coined for a particular occasion. If a term pops up in *Rolling Stone* and in the *Wall Street Journal*, one can conclude that it is not restricted to a small clique of speakers. (Landau 1984: 161)

Landau notes that there is a particular problem in determining whether a loan word like *barrio* has become "naturalized." The fact that Americans use it does not suffice to establish that its use is conventional in America (see §9.6). Landau faults standard citation practice for not recording whether the term is italicized or roman, which would help in making such decisions.

The neo-Gricean analysis avoids what I call the *explanation problem* with the Gricean analysis.[34] If an expression means μ, it is not merely a coincidence that people use it to mean the same thing, namely, μ. On the contrary, speakers mean μ because that is what the word means. As a contrast, suppose that by the wildest coincidence, a thousand speakers of English in the last decade had used the nonsense word "brong" in their private diaries, all of them using it to mean the same thing. It would not follow that "brong" means anything in English. Statements of the following form may therefore be both true and informative:

(11) The fact that people use e to mean μ is explained by the fact that e means μ.

I use the word "true" to mean "true," for example, because that is what the word means in English. The fact that "true" has the meaning it does is my reason for using it when I say things like "It is undeniable that (11) is true." The same thing holds for most people. So "people use 'true' to mean 'true'" holds in large part because "true" means "true." But if (6) told us what it is for an expression to mean something, (11) would contradict the irreflexivity of explanation. It would say: the fact that people use e to mean μ is explained by the fact that people use e to mean μ, which is absurd.

The result of coupling (11) with the neo-Gricean analysis (10) is quite different, saying that people use e to mean μ because it is conventional to do so. Collectively and individually, we do use the word "true" to mean "true" at least in part because it is conventional to do so. Our usage resulted from the convention, and the convention gives us a good reason for our usage. It is true in the same way that Americans drive on the right because it is conventional for Americans to drive on the right. The claim that something is conventional says more than that people do it. People make mistakes, get sick, and burp. But it is not conventional to do these things. A fortiori, we do not do them *because* it is conventional. A good definition of convention should entail that the performance of an action can be explained by the fact that it is conventional to perform it. Indeed, Definition 9.1 will entail that *the use of e to mean μ can be explained by a convention to use e to mean μ* (§9.3). So together with the

34 Cf. Alston 1965: 21, fn. 1; Sellars 1969: 112; Schiffer 1972: 13; Kempson 1975: 141; Bennett 1976: 8, 10; Biro 1979: 242; Yu 1979: 284; Platts 1979: 89–91; Harrison 1980: 167; Blackburn 1984: 90; Millikan 1984: 4–5; and Devitt & Sterelny 1987: 124–8. Contrast Smith et. al. 1952: 284.

neo-Gricean analysis, Definition 9.1 implies that the use of e to mean μ can be explained by the fact that e means μ. I think the neo-Gricean analysis is therefore equivalent to saying that e means μ iff people use e to mean μ because it is conventional to do so.[35]

The neo-Gricean analysis implies that many questions that might appear to be answerable by logic or simple introspection in fact require detailed empirical knowledge of the conventions of large groups of people,[36] which may not be readily available. Consider, for example, the vexed question of analyticity. As defined by Kant, a statement is analytic if the predicate concept is contained (properly or improperly) in the subject concept. Let us interpret "contained" in this definition as implying "being a part of" in addition to "being entailed by." In this sense, then, does "A brother is a male sibling" express an analytic truth? Given the grammatical structure of the sentence, the answer depends on whether the concept expressed by "male sibling" is a part (proper or improper) of the concept expressed by "brother." Since being a brother entails being a male sibling, and since it is hard for us to think "brother" without thinking "male sibling," the answer seems obvious. But these pieces of evidence from logic and introspection are not sufficient to answer our question. Given Definition 7.4, we need to know what "brother" and "male sibling" mean (in English). On the neo-Gricean analysis, that depends on what those terms are conventionally used to mean, which depends on what ideas they are conventionally used to directly express, by Definition 2.5. Neither introspection nor logic will suffice to tell us whether or not something is done conventionally, or by doing something else.

If that is not evident, consider this: being a sibling entails being a brother or a sister, and it is hard to think "sibling" without thinking "brother or sister." So if the former evidence proves that "brother" and "male sibling" express the same concept, then the latter evidence shows that "sibling" and "brother or sister" express the same concept. But these latter conclusions cannot both be right. The concept "sibling" cannot be a proper part of the concept "brother" if the concept "brother" is a proper part of the concept "sibling." So now the question becomes: when speakers use "brother" and "male sibling" conventionally, are they directly expressing the same idea? Or do they use the terms to express one idea directly ("brother" and "male brother or sister," respectively) and another indirectly ("male brother or sister" and "brother," respectively)?

35 Cf. Schiffer 1972: 120, 154.

36 Cf. Lance & O'Leary-Hawthorne 1997: §1.2.

The answer is not obvious, nor should it be. One piece of evidence, hardly conclusive, is that people generally learn the meaning of "brother" and "sister" long before they learn the meaning of "sibling." Another piece of evidence is that attempts at analysis have generally been unsuccessful for unstructured words in English and in other natural languages.[37] On the other hand, the hypothesis that the concept of a brother is the concept of a male sibling would provide a simple explanation of why "Bill is my brother" entails "Bill is male" and "Bill is my sibling."[38]

The interchange between Katz (1972; 1977b) and Putnam (1970b; 1973; 1975) over the analyticity of "Cats are animals" may be construed similarly. It is evident, I believe, that the word "cat" is conventionally used to express the idea of a feline animal. Since the idea of an animal is contained in that of a feline animal, "Feline animals are animals" is clearly analytic. It is a reasonable hypothesis, one that provides a simple explanation of a large body of data, that the word "cat" is used to express the idea of feline animal *directly*. That is, it is reasonable to assume that "cat" *means* "feline animal." This hypothesis explains, for example, why the restrictive clause in "Cats that are animals purr" seems funny. On the other hand, Putnam devised a famous thought experiment, involving a purely hypothetical case in which all of those particular entities that we had been calling "cats" (including the pet I am petting) turned out on inspection to be Martian robots, thus differing radically from dogs, rats, and all other animals. He in effect asked us to consider the following counterfactual conditionals:

If things had turned out that way, we would have discovered:
(a) that there never had been any cats, or:
(b) that cats really weren't animals.

Intuitively, most competent speakers of English (if we may extrapolate from philosophers) judge that (b) rather than (a) is correct. Replace "cats"

37 See §15.2 of this volume, and my *Nondescriptive Meaning and Reference*. Fodor (1975: 146–54) suggested that the speed with which we understand a sentence containing a term like "brother" favors the theory that it expresses a simple concept (see also Fodor et al. 1980: 502, §§2). But the data was insufficient to support this interpretation. If a definitional theory of understanding maintained that understanding a term involves (a) recalling its definition, and then (b) understanding the recalled definiens, then Fodor's prediction would be secured. But a more plausible theory is that understanding a defined term involves conceiving the concept that is expressed by its definiens, or coming to believe that it expresses that concept. No similar predictions can be drawn from this theory. Cf. Margolis & Laurence 1999: 17–18, who critically review other experiments on "processing load" that purport to show that the alleged complexity of lexical concepts has no "psychological reality."

38 See Katz & Fodor 1963; Katz 1964b; 1974.

with "feline animals" and the intuitive judgments are reversed. This evidence suggests that while the concept of an animal is central to our conception of cats, it is not part of the concept of a cat (§19.4). Although I believe Putnam is right, it is not my purpose here to reach a conclusion about this particular case. That would require consideration of a large body of linguistic data of the sort considered in the last few paragraphs. My point is that the controversy over the case supports what I am calling the neo-Gricean analysis. Given that analysis, it *should* be hard to tell whether "All A are B" is analytic when the concept of a B is fundamental to our conception of an A.

§8.6 THE USE/INTERPRETATION EQUIVALENCES

What I have called the neo-Gricean analysis focuses on the *production* of e, saying that e means μ provided that people conventionally *use e to mean μ*. But there is also a strong connection between word meaning and *interpretation*. Expression e means μ only if people are conventionally *taken to mean μ by e*.[39] Taking people to mean "female fox" by "vixen" is as customary as using "vixen" to mean "female fox." In general:

8.1 **Postulate:** *It is conventional for people to mean μ by e iff it is conventional for people to be taken to mean μ by e.*[40]

The claim that meaning μ implies being taken to mean μ is false: people are misinterpreted all too often. What Postulate 8.1 asserts is that a convention governing meaning implies a convention governing interpretation, and vice versa. This is not a necessary truth, although it is hard to imagine a case in which it fails. The equivalence is true in virtue of the contingent yet fundamental fact that *the common interest sustaining linguistic conventions is communication* (see §9.2). While there are exceptions to the rule, as we have emphasized in earlier chapters, speakers *normally* wish to communicate, and hearers normally want them to succeed. Speakers therefore need to use words in ways that will enable them to be understood, and hearers need to take speakers to mean what they mean. If it were conventional for people

39 Cf. Hungerland & Vick 1981: 69 on Hobbes, and Husserl 1900: 277, 302, 309. See also Laurence's (1996: 282–4) claim that the semantic properties of utterances are inherited from those of the mental representations that they are associated with in language processing.

40 Cf. Hockett's notion of "interchangeability," one of the "design features" he presumed to be universal and distinctive of human languages: "any speaker of a language is in principle also a hearer, and is theoretically capable of saying anything he is able to understand when someone says it" (1958: 578).

to use "vixen" to mean "female fox" but not for people to take others to mean "female fox" by "vixen," people would fail to communicate, and the convention would not sustain itself.

Taking S to mean μ by e is to be interpreted as meaning *thinking of S as meaning μ by e*. H may think of S as meaning μ in a number of ways. Normally, H *knows* and *recognizes* that S means μ, or at least *believes* that S means μ. That is, taking is normally understanding. But in other cases, H may just *assume* or *hypothesize* that S means μ. This form of taking occurs, notably, when we are unsure what a speaker means, and are trying out various interpretations. A student of Heidegger, for example, might say "Let's try taking him to mean _____." If that interpretation makes good sense of Heidegger, the student might be led to conclude that that was what Heidegger meant. If Heidegger does not make sense on that interpretation, the student might try another. Whether H knows that S means μ, or just hypothesizes that he does, H is *interpreting* S as meaning μ.

Taking may also be partly perceptual. Normally, we *hear* the speaker *as* meaning μ by e. This involves more than simply hearing the speaker or his words. One can see a Necker cube without seeing it as coming out of the blackboard, and one can hear the clarinet in *Peter and the Wolf* without hearing it as the duck. To hear the speaker as meaning μ, the idea that he means μ must be occurring to the speaker, and the occurrence of that idea must be connected in the right way with the hearer's auditory sense impression of the speaker. Believing or knowing that the speaker means μ is not required for hearing him as meaning μ, as the duck and cube examples also illustrate.[41]

Note that while knowing or believing that S means μ by e is not subject to volition, taking S to mean μ by e is, in the sense defined. The voluntary character of interpretation is especially striking when the speaker says something ambiguous, such as "The car is hot." Then we can "flip" from one interpretation of "hot" to another, at will. We may be able to interpret the speaker as meaning that the car's temperature is high even though we do not know what he means, and are unable to form an opinion.[42] *Taking* is a voluntary action even though in some cases we cannot help taking things a certain way, just as *breathing* is voluntary even though people sometimes breathe involuntarily.

41 Cf. Laurence's arguments for the "modularity" of the language processor (1996: 282–96).

42 Cf. Bennett 1976: 178–9 and Lewis 1975. Since both viewed the speaker as trying to produce belief, both viewed the hearer's response as believing. For this reason, both were led to deny that conventions must be voluntary actions.

In light of the equivalence between conventional use and interpretation, the neo-Gricean analysis could be strengthened as follows:

(12) Expression e means μ iff people conventionally use e to mean μ and are conventionally taken to mean μ by e.

The truth of Postulate 8.1 ensures that the strong neo-Gricean analysis (12) will be true if and only if the weak one (10) is. Which analysis holds if Postulate 8.1 is imagined to be false? Such cases are far-fetched, since people would regularly fail to communicate in virtue of their conventions. But if we consider such a possibility, it seems to me that expression meaning would go along with use rather than interpretation. If people conventionally use "red" to mean "red," then "red" means "red," even if people conventionally take others to mean "blue" by "red." If a speaker means "red" in conformity to the convention, then that is what "red" means on that occasion. And if the word regularly means "red," then that is what the word means *simpliciter*.

Together with the neo-Gricean analysis, Postulate 8.1 provides part of the conceptual basis for the use of *linguistic intuitions* as evidence in semantics.[43] The other part is provided by a companion postulate relating the interpretation of speakers to the interpretation of words. Speakers of English not only take *others* to mean "female fox" by "vixen," they take *the word* "vixen" to mean "female fox." Since the former is habitual for competent speakers, so is the latter.

8.2 **Postulate:** *People conventionally take others to mean μ by e iff people conventionally take e to mean μ.*

The truth of Postulate 8.2 is rooted in the contingent but well-established fact that speakers of English have the concept of word meaning, and know at least implicitly how word meaning is connected to speaker meaning. Postulate 8.2 may be regarded as the kernel of truth in the naive analysis of speaker meaning, according to which S means μ iff S used an expression meaning μ (see (4) of §2.1).

Postulates 8.1 and 8.2 obviously imply that *e is conventionally used to mean μ iff e is conventionally taken to mean μ*, which directly relates use to interpretation of a different sort. Given these postulates, the neo-Gricean analysis is equivalent to:

(13) Expression e means μ iff e is conventionally taken to mean μ.

43 Cf. Katz & Fodor 1963; Fodor & Katz 1964: Chapter 1, esp. fn. 18; Katz 1964b; Chomsky 1964: 79–81; 1965: 20–1; Chierchia & McConnell-Ginet 1990: 16.

This consequence of the neo-Gricean analysis cannot be regarded as an analysis of the concept of meaning, however, without circularity. It would be defining word meaning in terms of word meaning. The right side of (13) may nevertheless describe a convention that is empirically necessary and sufficient for a word to have a meaning.

The connection we have explored between meaning and interpretation may also be behind the doctrine that a theory of meaning is a theory of interpretation.[44] However, interpretation is a cognitive process. It is something that people do when they perceive words. Word meaning is not a cognitive process, and not something that people do when they perceive words. So a theory of meaning cannot literally be a theory of interpretation. The latter would be an exercise in cognitive psychology. The most we can say is that words have the meanings they do because they are used and interpreted in certain ways.

§8.7 THE EXPRESSION/COMMUNICATION EQUIVALENCE

The neo-Gricean analysis defines word meaning in terms of speaker meaning or expression. I argued earlier that speaker meaning or expression must be defined in terms of the speaker's intention to produce an indication of his own thoughts or ideas, rather than the speaker's intention to produce thoughts or ideas in others (see Chapters 2–4). For example, a speaker who insults someone in Rumanian, knowing full well that the hearer does not understand Rumanian, may well mean "You are a filthy swine" even though he does not intend, want, or expect to produce that thought in the hearer. At the level of word meaning, however, it is just as conventional to use words to communicate thoughts as it is to use words to express thoughts. Indeed, the following equivalence appears to be universally true. Let "μ_i" stand for terms expressing i and used ideo-reflexively, as in §7.6.

8.3 **Postulate:** *It is conventional for people to use e to mean μ_i and express i iff it is conventional for people to use e to communicate i and produce it in others.*

Given this connection between expression conventions and communication conventions, the neo-Gricean analysis of word meaning can be strengthened in another way.

44 Cf. Dummett 1975; 1976; Lycan 1984; 9; Schiffer 1987a: 123, 214.

(14) Expression e means μ_i iff people conventionally use e to express, produce, and communicate i.

The contingent truth of Postulate 8.3 ensures that the strong neo-Gricean analysis (14) will be true if and only if the weak one (10) is.

While it is easy to recall actual cases in which speakers mean something without intending to communicate, I do not believe that there are any cases in which an expression means something even though it is not conventionally used to communicate. I cannot think of any actual exceptions to Postulate 8.3. Hypothetical counterexamples can be imagined without contradiction or incoherence, but are inevitably far-fetched and unrealistic. That is, while Postulate 8.3 is logically contingent, it is by no means a mere accidental generalization. For it is hard to imagine how the use of a word to express a certain idea could become conventional in a population if the use of that word did not evoke the same idea in others. The truth of Postulate 8.3 is a consequence, I believe, of two important facts. First, communication is the common interest that sustains meaning conventions (Chapter 9). Second, thoughts and ideas are special in that their communication implies transmission (see §5.5).

§8.8 THE EXPRESSION/INDICATION EQUIVALENCE

We observed in Chapter 3 that expressing an idea, and meaning something, requires not only that the speaker provide an indication that the idea is occurring to him, but also that he do so without covertly simulating an unintentional indication. If Stu contorts his face and spits out some food in a careful attempt to fake a reflexive response to awful food, the covert nature of his simulation makes it false that he meant "the food is bad," even though he did intend to provide an indication that he was thinking that thought. By their very nature, it seems, *covert* simulations could not become *conventional* indications of an idea. If it were conventional for people to perform a particular act $\boldsymbol{\sigma}$ that simulates an unintentional indication of idea μ, then surely people would generally recognize that $\boldsymbol{\sigma}$ is not a genuine unintentional indication. In that case, it would be unusual for speakers to *intend* $\boldsymbol{\sigma}$ to be incorrectly taken as genuine, for it would be irrational for them to expect others to be fooled. So it could hardly be conventional for speakers to provide an indication of their ideas by covertly simulating an unintentional indication. The same holds whether the indication is intended to be direct or indirect.

We also observed in Chapter 3 that an action *intended* to be an indication of something need not be an actual indication. But if it is *conventional* for people to produce an expression e with the intention of providing an indication that idea μ is occurring to them, then the use of e will *be* an indication of that idea. The production of e will conventionally be accompanied by the occurrence of idea μ. Tokens of e will commonly be caused by mental events containing occurrences of the idea, and, in a feedback loop, perception of the tokens will activate further occurrences. Because of these connections, the use of e provides evidence of idea μ, and observation of e gives hearers reason to believe that the idea μ is occurring to the speaker. An additional postulate is plausible, therefore, that links conventional expression and indication.

8.4 **Postulate:** *It is conventional for people to use e to (directly) express i iff the use of e is a conventional (direct) indication of i.*

While speaker expression and meaning *simpliciter* cannot be equated to the indication of ideas (or other mental states), *conventional* expression and meaning is coextensive with the conventional indication of ideas (or other mental states).

Traditional ideational theorists like Locke and Hobbes explained what words mean in terms of what utterances of the words signify in the evidential sense that concerns indication. "Dog" means "dog," they observed, in virtue of the fact that utterance of the word "dog" is a "conventional sign" of the occurrence in the speaker of the idea "dog." Their observation is now a simple corollary of Postulate 8.4 and the neo-Gricean analysis (10).

The basic neo-Gricean analysis has two major deficiencies. It does not account for either the productivity of word meaning, or its relativity to languages. These problems will be addressed in Chapters 10 and 11. We will show there how the case of word meaning for which the neo-Gricean analysis does hold provides the basis for all word meaning. Chapter 9 will analyze the concept of convention used in the analysis, and show how the neo-Gricean analysis can differentiate implicature conventions that give rise to meaning from those that do not.

9

Conventions

To fully understand the thesis that basic word meaning is conventional speaker meaning or expression, and to avoid misguided criticism, we need to know what conventions are. Briefly, we may say that "convention" should be understood as denoting *arbitrary social practices* or *customs*. Constituting standards of correct usage, they are one type of *rule*. This chapter will be devoted to clarifying what conventions in this sense do and do not entail. We will develop Lewis's idea that conventions are regularities in action that are socially useful, self-perpetuating, and arbitrary. There is no requirement, on our definitions, that the regularities be nearly universal or mutually known. And of course there is no requirement that conventions result from agreements. We will take some pains to explain how word usage can be objectively correct or incorrect if it is arbitrary and conventional. The fact that conventional regularities may have exceptions allows languages to change over time, and the self-perpetuating character of conventions coupled with linguistic variation leads to evolving families of languages. The same facts make it difficult to assign precise boundaries to languages, as we will see in Chapter 11.

§9.1 DEFINITION

It is often said that conventions are *agreements*.[1] Indeed, in one sense, the word "convention" denotes an international agreement, and in another

1 The locus classicus is Hermogenes in Plato's *Cratylus* (383b). See also Reid 1764: 32; Whorf 1956: 213; and Schein & Stewart 1996: viii. Hume (1739: 490) calls conventions "agreements," while denying that they are "promises." Gilbert (1996: 110) describes conventions as "quasi-agreements," saying that "it is as if they agreed."

denotes formal meetings designed to secure agreements. But in the sense we are concerned with, most conventions, including linguistic conventions, are not and did not result from agreements. It is hard even to imagine the first language arising by agreement. How could people agree to such a thing without using a common language? In the inimitable words of Bertrand Russell, "We can hardly suppose a parliament of hitherto speechless elders meeting together and agreeing to call a cow a cow and a wolf a wolf."[2] Linguistic conventions do occasionally arise from agreements, as when scientific congresses succeed in standardizing terminology. But other origins are much more common, as when a metaphor dies or a stipulative definition catches on. Thus Edward Kasner coined the wood "googol"; his readers used it because he did; and so on and on. In some cases, agreements evolve into conventions that undermine the purpose of the agreement, as when brand names become generic names.

Bach and Harnish (1979: 108–10) have defined conventions as "count-as rules," actions that, if done in certain situations, count as doing something else.[3] The rule whereby baseball players score a run by crossing home plate is such a rule, since crossing home plate counts as scoring in certain circumstances. Similarly, saying "I promise" counts as promising in certain circumstances. However, the notion of a count-as rule does not fit two of the paradigm examples of conventions that Bach and Harnish themselves cite. "Money is a conventional means of exchange," they say. Furthermore, "It is a commonplace, however unexplicated, that language is a system of conventional means for communicating." It is conventional to give the grocer money in exchange for food. But giving the grocer money does not count as making an exchange. What counts as making an exchange is giving the grocer money in exchange for food. But the connection between making an exchange and giving the grocer money in exchange for food is not conventional in any sense, for the latter necessarily entails the former. Similarly, we often communicate by uttering sentences; but uttering sentences never counts as communicating. What

2 Russell 1921: 190. See also Reid 1764: 4.2, 5.3, 6.24; Quine 1936; 1969b; Alston 1964a: 57; Lewis 1969: 2, 83–8; Bennett 1976: 206–10; Ullmann-Margalit 1977: 74–6; Grice 1982; Blackburn 1984: 119; Suppes 1986: 113; and Gilbert 1996: 253–61. Contrast Schein & Stewart 1996: xiii. Jamieson (1975) misguidedly argues that Lewis's analysis fails because it does not fit all of the Geneva Conventions relating to war. Kretzmann (1967) reports that while the great associationist philosophers did not take the conventionality of language to imply its origins in agreement, their philosophy of language was obscured from antiquity onward by confusion on this point.

3 Cf. Strawson 1964: 292; Searle 1969: 33.

counts as communicating is uttering sentences with certain intentions that are recognized by an audience. This connection holds in virtue of the nature of communication, not because of any arbitrary convention. The rules in virtue of which words have meaning are not count-as rules, either. Using the word "dog" does not count as meaning "dog." What counts as meaning "dog" is using the word "dog" with certain intentions. But the connection between using "dog" with those intentions and meaning "dog" is necessary rather than conventional.

I believe that the most accurate definition of convention was suggested by David Lewis in an informal characterization that echoes a famous passage in Hume (1739: 490).

> Conventions are regularities in action, or in action and belief, which are arbitrary but perpetuate themselves because they serve some sort of common interest. Past conformity breeds future conformity because it gives one a reason to go on conforming; but there is some alternative regularity which could have served instead, and would have perpetuated itself in the same way if only it had got started. (Lewis 1975: 4–5)

This characterization perfectly fits Lewis's paradigm example of driving on the right. It also fits linguistic conventions admirably. Lewis's official theory however, offered initially in 1969 and revised in 1975, was much more complex, requiring "mutual knowledge," a highly specific game-theoretic preference structure, and near-universal regularities.[4] These requirements are not satisfied in the case of linguistic or other conventions, as we shall see later. So if Lewis's official definitions were used, the Gricean thesis that word meaning is conventional speaker meaning would be false. We will adopt a modification of Lewis's informal gloss as our definition.

9.1 **Definition:** *A convention is a regularity that is socially useful, self-perpetuating, and arbitrary.*

In the rest of this chapter we will explain the key terms in Definition 9.1, show that the stronger requirements of Lewis's official definitions are too strong, and explore the way in which convention is the proper standard of correct usage.

4 See also Schelling 1960: Chapters 3–4; Grice 1968: 233; Schiffer 1972: Chapter 5; Burge 1975; Bennett 1976: Chapter 7; Grandy 1977; Ullmann-Margalit 1977: Chapter 3, esp. p. 96; Pateman 1982; Blackburn 1984: 82–92, 118–22; S. Miller 1992; Gilbert 1996; Laurence 1996: 273, 284–5; Avramides 1997: §10.

§9.2 SOCIAL UTILITY

In order for an action to be *socially useful*, it is necessary but not sufficient that it serve the interests of individual members of the group. The interest served must be *mutual* or *collective* as well as *common*, something that people want not only for themselves but also for others or for society as a whole. In the clearest cases, individuals are motivated to coordinate their actions for a common goal. Thus people regularly drive on the right in America because drivers have a common and mutual interest in avoiding head-on collisions. Languages are socially useful because they serve a common, mutual, and collective interest in *communication*. That is the sociobiological function of language. Because people individually and collectively wish to communicate, they need to coordinate their actions. Speakers need to use a language that their audience will understand. Because people understand the language they use, speakers are generally better off using the language that their audience uses. It is not nearly as important which language speakers and hearers use, as long as it is the same language.

In earlier chapters, we have argued that Grice and his followers erred in trying to equate speaker meaning with attempted communication, while granting that there are important connections between the concepts of meaning and communication. We can now formulate an indirect but nonetheless important link: *the desire and need to communicate is the common interest that generates conventions to use words to mean things, which determine what the words mean.*

In some cases, a convention has a clear social function even though it is not clear that there is a common *goal* among participants.[5] It is conventional for men in many countries to wear ties at business functions and formal occasions, for example. Wearing ties is socially useful, even though most of us are hard pressed to think of any purpose that it serves. It is not imperative that we clarify what social utility amounts to in such cases, because linguistic regularities serve a clear and vital purpose.

§9.3 SELF-PERPETUATION

Conventions are *self-perpetuating* in special ways. We will identify six mechanisms by which social practices are reinforced and sustained: *precedent, association, habit, enculturation, normative force,* and *social pressure*.

5 This issue was raised by Thomas Bartelborth.

(1) *Precedent*. Given the common interest served by a convention, the fact that people have conformed to the regularity in the past gives people a good reason to continue conforming to it. Previous conformity thus serves as a precedent. Sometimes the precedent works directly: the fact that Americans have driven on the right before gives me an excellent reason to drive on the right now. Sometimes, the precedent operates indirectly. Steering wheels are conventionally installed on the left side of cars made in America. This is done in part because people drive on the right, it being easier to drive on the right when steering from the left. Yet for the same reason, the practice of driving on the right has flourished in part because steering wheels have been installed on the left. Moreover, steering wheels are installed on the left principally because drivers in America are used to having them on the left, and plants are equipped to produce cars with left-sided steering. Yet people are so habituated, and plants so equipped, because steering wheels have been installed on the left before. So, indirectly, steering wheels are installed on the left today because they were installed on the left in the past.

In the linguistic case, precedent works both directly and indirectly. The very fact that Americans have used English before is a good reason for Americans to use English now. It is easier to communicate in a community with one language than in one with many languages. More importantly, the fact that Americans have used English before gives Americans a reason to expect that their audience will understand English now. In both ways, the fact that others before me have used the word "true" to mean "true" gives me a good reason to do so now. The reasons provided are seldom *conclusive* in any sense, but they do lead Americans to continue to use English.

Dictionaries represent another indirect way in which precedent operates. Lexicographers base their entries on citations of past usage. Current users consult dictionaries as authorities on current usage. Most of the time, the fact that a usage is described in a dictionary gives people a compelling reason to adopt it. Usage manuals have the same effect when their advice is based on intimate knowledge of exemplary literature. Written language changes more slowly than spoken language, because the permanence of written documents means that precedents can operate over a long period, whereas spoken precedents last only as long as our memories (Daniels & Bright 1996: 1).

Lewis wanted to emphasize the *rationality* of conventions, which distinguishes them from mere de facto rules, such as the rule that speakers regularly punctuate their speech with "uh," and nearly always misspell

some words in a long document. It is not a mere accidental generalization that speakers of English use "red" to mean "red." The regularity exists because each speaker rationally strives to do his or her part to make it exist. A speaker's actions are motivated and justified by his or her preference for the expected outcome of conformity versus nonconformity. The "performance errors" that occur when people misspeak as a result of inattention, memory limitations, drunkenness, or fatigue thus create no problem for the neo-Gricean thesis that meaning is determined by conventional use. For in such cases, the speaker is not actually following the precedent that he intended to follow.

I believe that Lewis's concern with rationality explains why he placed exclusive emphasis on precedent as the mechanism of self-perpetuation. However, the fact that linguistic precedents provide a *genuinely good* reason to follow them is immaterial for our purposes. What matters is that the reasons are *regarded as good*, and consequently *motivate* speakers to perpetuate the convention. Moreover, several other mechanisms are capable of sustaining and reinforcing conventions, and are no less important in practice.[6]

(2) *Association*. The regular use of a word to express an idea results in a mental association between the word and the thing meant, so that images and perceptions of the word call up the idea of the thing, and vice versa (see Chapter 18). The more often the two are paired, the stronger their association. The association between the word and a meaning together with knowledge of the precedent connecting them commonly makes that use of the word "*salient*," and thus more likely to be noticed as an alternative and selected.[7]

(3) *Habit*. Regular action leads to a habit in each individual, which is reinforced by subsequent actions done out of that habit. At an early age, English speakers develop the habit of using the word "red" to mean "red," and "dog" to mean "dog." In the case of linguistic conventions, there are two sets of reinforcing habits. The fact that speakers regularly use e to mean μ leads hearers to automatically hear e as meaning μ (§8.6); the more frequent the usage, the more automatic the understanding. The fact that audiences habitually interpret e as meaning μ normally gives speakers a good reason to use e to mean μ; the more automatic the understanding, the better the reason for use. Linguistic habits are often strong enough to interfere when the speaker has consciously decided that it would be

6 Cf. Jamieson 1975.

7 Cf. Schelling 1960: 57, 67, 91; Lewis 1969: 35–6; Schiffer 1972: 145–8; Ullmann-Margalit 1977: 83–4; and Gilbert 1996: Chapter 1.

best to use another language or terminology. The fact that an individual habitually conforms to such a complex set of regularities in a flexible, creative, and purposive manner implies that the individual has some sort of knowledge or internal representation of the regularities, as Chomsky (e.g., 1965: Chapter 1) and his followers have stressed. Other evidence suggests that linguistic habits have a specialized neural basis (e.g., Laurence 1996: 284–92).

(4) *Enculturation*. Language is passed on as a *tradition* in the process of enculturation.[8] Because adult Americans use English, American children learn English. Children learn it at home, are taught it in school, and pick it up from their peers. Adults who join the group learn the conventions from older members. The transmission of a language during child rearing is facilitated by features not found in other traditions. For example, the set of speech sounds that an individual is capable of perceiving and producing is influenced by the set of speech sounds that he or she heard as a child.

The use of one word rather than another to mean something is often a conscious, reflective process. So known precedents are readily able to function as reasons. But conscious decisions to follow precedents appear to play only a small role in linguistic enculturation. For example, Americans who spend a year or two at Oxford often pick up a British accent. But they do not actively try to change their accent, and typically do not realize that it has happened. It does not seem accurate to say that they decided to adopt British pronunciations, for the reason that others were speaking that way. Similarly, when speakers of nonstandard dialects interact with those of the dominant dialect, the grammar of the former gradually becomes standardized through a process that seems much more subconscious than imitation of a selected precedent (Labov 1970: Chapter 2). Labov observed that most linguistic rules acquired by the native speaker are not consciously recognized and never violated, such as the complex contraction rule that allows "He's here" but not "Here he's." Finally, the Creoles spoken natively by children whose parents speak a pidgin language have been found to be more complex and systematic than the pidgin the children hear their parents speak (Rickford 1992: 226–7; Laurence 1996: 291–2).

It is not entirely clear whether a practice that is not transmitted from one generation to the next counts as a convention or a custom. Are fashions and fads short-term conventions or just conventions in the making? It is similarly unclear whether linguistic fads and fashions count as part of the language. If a slang word created by today's teenagers dies out when

8 Cf. Hockett (1958: 579–80; 1966: 11) and Asher (1994: 876).

they reach adulthood, do we count the word as part of English during its brief period of use? Was its slang use conventional while it flourished? I am inclined to answer "yes" to both questions, as long as the practices are self-perpetuating by the other mechanisms. So I conclude that traditional transmission is not essential for conventions or for linguistic meaning, although it is the norm. A more important conclusion, however, is that if meaning is determined by convention, then these issues *should* be unclear. The concept of word meaning is inherently vague in ways in which the concept of speaker meaning is not, and the Gricean analysis correctly identifies the source.

(5) *Normative Force*. Conventions also serve as *generally accepted standards of correctness*.[9] The left side of the road is considered the right side to drive on in England, the wrong side in America. People often misspell "receive," putting the "i" before the "e." Some do so habitually. But few who misspell the word do so because they want to conform to the regularity of spelling it with "i" before "e." On the contrary, people regard this spelling as a mistake. They correct themselves (and others) whenever they can. The same goes for conventions of meaning. When I find, as I often do, that I have used "their" to mean "there," I do not regard my action as just unusual or accidental; I consider it to be a mistake. I judge my usage to be incorrect, improper, and unacceptable. The more closely I monitor and control my writing, the less often I make the mistake. In general, people are often proud when they speak correctly, and embarrassed when they do not. Conventional regularities are thus generally accepted *rules* or *norms*, by which agents judge and guide their actions. Conventions are more than statistical regularities: they are de jure rules "in force" in the community.

(6) *Social Pressure*. Conventions are partly sustained by social pressure. People use conventional norms not only to guide their own behavior, but also to criticize or correct the behavior of others. Children are praised and otherwise rewarded for learning to speak and write properly. Adults who do not speak correctly may be shunned or denied jobs. Change is often greeted negatively as corruption – until, of course, the change becomes fully conventional. Editors alter manuscripts, sometimes against the

9 Cf. Quintilian, *Institutio Oratoria*: 72–3; Locke 1690:§3.2.8, Priestley 1762: 136–9, 184; Smith et al. 1952: Chapter 12; Leonard 1929: Chapters 1–2, 9–10; Fries 1940: Chapter 1; Alston 1964a: 41–4; 1964b: 57–8; 1974: 19, 25; Lewis 1969: 97–107; Rosenberg 1974: 43–5; Ullmann-Margalit 1977: 12–13; 85–93, 96; Kripke 1982: 89–95; Blackburn 1984: 83; Landau 1984: Chapter 5; Fasold 1990: Chapters 7–8; O'Grady et al. 1993: 12; Gilbert 1996: 78–86, 107–12; Lance & O'Leary-Hawthorne 1997: Chapter 3; and Yagisawa 1998: 449. Contrast Ziff 1960: 30–1, 34–8.

author's will, to make them conform to conventional usage. Manuscripts departing too much from convention are simply rejected.

Because conventional actions are regularities that perpetuate themselves in the six ways we have discussed, *the fact that an action is conventional can explain why people perform it*. People drive on the right today because it is conventional to do so. They drive on the right today because the fact that people have done so before gives them a good reason to do so today, because they have been taught to drive on the right, and so on. In the same way, people use the word "red" to mean "red" today because it is conventional to do so. Normative forces and social pressures, along with the force of precedent, association, and habit, lead people to use "red" to mean "red" today. Thus the self-perpetuation clause of Definition 9.1 is what enables the neo-Gricean analysis of word meaning as conventional speaker meaning (§8.5) to avoid the explanation objection, and to secure the result that people use "red" to mean "red" because "red" means "red."

The statement that a regularity is conventional entails that it maintains itself in certain ways, but entails nothing about how it originated. In particular, the statement does not entail that the regularity originated in an agreement of any sort. The definition of convention also leaves open the question of whether conventions are laws: they may be, but need not be, codified or reinforced by legal sanctions. The convention of driving on the right does have the force of the law behind it, but the convention of starting a letter with "Dear ____," does not. It is undoubtedly important to make driving on the left illegal, and to punish infractions, because even one violation can have fatal consequences. But the convention would surely remain quite robust without legal reinforcement.

§9.4 ARBITRARINESS

Conventions are *arbitrary* in the sense that there is another possible regularity in action that could have served the same common interest, and would have perpetuated itself in the same ways if only it had got started. The arbitrariness of symbols stands in marked contrast to the statistical or causal character of indexes and the representational character of icons.[10]

10 Cf. Locke 1690: §3.2.1; §3.2.4; Peirce 1931–5: vol. 2; Ziff 1960: 25–6, 57–8; Alston 1964a: Chapter 3; 1974: 19, 25; Bennett 1976: 14, 149; and Anttila 1989: 12. According to Kretzmann, "Sources very late in antiquity attributed to Pythagoras (fl. 530 B.C.) the view that although the soul assigned names to things, it did so not arbitrarily but on the basis of a natural connection between them, somehow like that between mental images and their originals" (1967: 359).

Conventional regularities are not laws of nature, but originate in "the artifice and contrivance of men" (Hume 1739: 491). Some features of a language have no alternative. The fact that expressions are perceivable is an obvious example. If the Chomskyan thesis of an innate language acquisition device is correct, there will be many unobvious examples. The necessary features of language are not restrictive enough, however, to make only one language humanly usable, as the thousands of languages spoken on Earth today prove. There is nothing in the nature of the human mind, the word "hood," or the idea of an engine cover in virtue of which one expresses the other. Hence "hood" could just as well mean "tire," and the idea of an engine cover could just as well be expressed by "hatch" or "torp." The American interest in communication would have been equally well served by Spanish. If Americans spoke Spanish, that would give Americans a good reason to continue speaking Spanish. American children would learn Spanish because adults use it. And Americans would criticize each other for speaking Spanish incorrectly.

Nonarbitrary practices are not conventions. For example, it was the practice of most city dwellers during the Middle Ages to dispose of their sewage in the street; it is our practice not to do this. The medieval practice may have been a convention, although I suspect that there were no practical alternatives. But we have compelling reasons today, both medical and aesthetic, for refraining from the medieval practice. The fact that others refrain from disposing of their sewage in public places is not one of them. So our practice is not properly called a convention. Similarly, there is a sense in which my car has a conventional gasoline engine. But in the sense that we are interested in, it is not a mere convention that gasoline engines are installed in nearly all automobiles today. Given today's technology, no other type of engine can achieve the same performance for the price (or so most people believe).

Sometimes the arbitrariness of conventions is obscured by the fact that they are not arbitrary *given* the existence of other conventions. For example, given that people drive on the right in America, it is not arbitrary that steering wheels are placed on the left. Steering from the left makes it significantly easier to drive on the right. But the placement of steering wheels is nevertheless arbitrary in the relevant, absolute sense. For both practices could be altered in tandem and, with sufficient time for retooling, would serve the same social interest equally well, perpetuating themselves as surely. In the linguistic case, the practice of using English in America is not arbitrary *given* that Americans conventionally interpret other Americans as speaking English. But both practices could be altered

and, with sufficient time for reeducation, would serve the purpose of communication equally well, perpetuating themselves in the same way. The use of one's native language is also nonarbitrary given the habits, associations, enculturation mechanisms, and social pressures sustaining the conventions of the language. People are capable of learning another language. They can fairly easily make a few terminological innovations. But until they learn a new language, they have little choice but to use what is basically their native tongue. Thus the fact that a man is alone usually does not liberate him from the need to conform to the conventions of his language community.

Sometimes the arbitrariness of a convention is obscured by its normativeness. Gilbert (1996: 69–70) has observed, for example, that we do not ordinarily think of the custom of sending thank-you notes after attending a dinner party or receiving a birthday present as arbitrary. Without reflection, we think of it as altogether right and proper to do this. But on reflection, it is clear that there are alternative ways of expressing gratitude, which could perpetuate themselves in the same way. We could instead adopt the practice of sending small gifts as tokens of appreciation, or flowers; we could make donations to charity in the host's name, or perform some personal service; we might even sing a "thank-you song" for the host. Note that it is only the way in which we express gratitude that is arbitrary. There is no substitute for the act of expressing gratitude. Similarly, it is our manner of communicating that is arbitrary, not the act of communication itself.

Following the ancient Greeks, some have concluded that onomatopoetic words and many sign language signs are "natural" rather than "arbitrary" because they are "iconic."[11] This statement is true only if the terms "natural," "arbitrary," and "iconic" are understood in a way quite different from the way in which we've used them. First, "quack" does not usually represent the sound of a duck iconically in the way that a road map represents a road system iconically. The word is iconic in Peirce's sense only if it is used imitatively, as when we say "The duck went *quaaaack, quack, QUACK*." Here we are intended to infer properties of the duck's vocalization from nonlinguistic properties of our utterance. We are supposed to infer that the duck's first quack was longer than the

11 Stout 1899: §§5.5–5.7; Hockett 1958: 577–8, 1963: 10; Deuchar 1984: 16–17; Kyle & Woll 1985: 113; Asher 1994: 876; Schein & Stewart 1996: ix, 32, 48; Robins et al. 1997: 556. Compare and contrast Saussure 1916: 67–70; Hockett 1958: 298–9; Ziff 1960: 25–6; Lyons 1971: 4–5; and Anttila 1989: 12–4.

second and softer than the third because the speaker's first utterance of the word "quack" was longer than the second and softer than the third. But such a use is very special, even for onomatopoetic words. "Quack" is not an icon in "The duck quacked three times, once long, once short, and once loud." Neither the iconic use nor the symbolic use constitutes a natural sign in the evidential or black cloud sense, unless the speaker happens to be a reliable source of information about the duck.

Second, the fact that "quack" represents the sound of a duck is nonarbitrary only in the sense that there is a reason making the word *appropriate* to its meaning. It is common in English and other languages to use onomatopoetic words for animal vocalizations, and the sound "quack" is reasonably similar to a duck vocalization. Almost no words are arbitrary in the sense of there being no reason that makes them appropriate. Thus "fork" used to describe a road, being a dead metaphor, is appropriate because of a similarity between the prior denotation of the word and the road configuration. Noun compounds like "airplane" and "nose count" illustrate other word formation principles, as do Latin- and Greek-based words like "hematite" and "metronome." Even a made-up word like "Xerox" was introduced as a trade name in part because of its similarity to "Kodak."[12] Nevertheless, the use of "quack" to mean the sound a duck makes is arbitrary in the sense relevant to convention: many other words could have been used to mean that sound. "Yack" would have worked sufficiently well in either the imitative or the denotative use. In fact, the French use "coin coin" (pronounced *kwɛ̃ kwɛ̃* with nasalized ɛ̃) to denote the sound of a duck – and also the sound of a motor horn.[13] Furthermore, the denotative use does not have to be onomatopoetic. The word "peal" for the sound of a bell (*carillon* in French) contrasts with "dong" in not being imitative. "Peal" derives from the Middle English word meaning "strike" or "beat." Indeed, most words for human vocalizations are not onomatopoetic. Consider "yell," "scream," "sing," "whistle," "whimper," "whisper," "cry," and of course "speak." "Hum" is exceptional.

12 In *Nondescriptive Meaning and Reference*, I observe that in addition to meaning conventions, which govern the use of words, there are naming or word-formation conventions, which govern the introduction of words.

13 Catherine Ball of the Georgetown University Linguistics Department has collected onomatopoetic dog sounds from over thirty-five languages. Nearly all are different, and the exceptions (Hebrew and Arabic, Chinese and Korean) may well have had a common origin. The range is quite large, from *bow wow* in English to *wan wan* in Japanese, *gav* in Greek, *ham ham* in Albanian, and *bup bup* in Catalan. See: www.georgetown.edu/cball/animals/dog.html

For our purposes, a *regularity* should be defined as *a common way of doing things*. "Dog" is regularly used to express the idea of a dog in America. The use of "dog" to express the idea of a dog is both a common way of using the word and the common way of expressing the idea. If W is the conventional way of doing A, then W is "the way A is done." Definition 9.1 can therefore be reformulated as follows: *A conventional way of doing things is one that is common, socially useful, self-perpetuating, and arbitrary*. And the neo-Gricean analysis can be rewritten to say that e means μ iff the use of e to mean μ meets this description.

A conventional regularity may also be described as a *rule*, one that is in force in a community. Not all rules are conventions, however. Since a convention in our sense is a common way of doing something, an action that has never been performed cannot possibly be conventional. Schiffer (1972: 150) denied this, observing that "there may be conventions governing behaviour in case of enemy attack which have never been acted upon because there has never been an enemy attack." "Rule" is appropriate here, or "agreement," but not "convention" in the sense that we are concerned with. We cannot say that it is our practice to launch a massive counterattack in case of enemy attack, if we have never been attacked, nor that it is conventional to do this. There may be a rule specifying such an action, to which the relevant decision makers have agreed. But the action does not count as a convention, custom, or practice until it perpetuates itself in the manner just specified. Gilbert has recently revived Schiffer's argument. "Suppose that you and I agree that if today's battle is lost you will wave a black flag from the hilltop. . . . It seems natural to say that we have, in making this agreement, adopted a certain 'signaling convention.'" (1996: 93). Again, "rule" or "agreement" is appropriate. But we cannot yet say that it is conventional for us to use that signal, nor that it is our practice to wave the flag to signal defeat. We cannot say that until the action becomes common.

The subtle difference that I am noting makes a difference for our project. Suppose that a group of English speakers agree to use "red" to mean "sloop." They have, in the sense that Gilbert had in mind, adopted a convention. But their action does not suffice to make "red" mean "sloop." That will not happen until it is conventional for English speakers to use "red" to mean "sloop," in the sense of "convention" I have been trying to clarify. The group would have to begin using "red" as they agreed, and

their usage would have to perpetuate itself and spread sufficiently through the population. This is necessary, I believe, even if, *per impossible*, the stipulating group included all one billion or so speakers of English. The English language would not change until the agreement among English speakers came to be reflected in their practice.

An action *is* a convention only if it is common *today*, meaning not "this day" but "these days." A regularity over the recent past suffices. It is conventional in America today to use the word "decagon" to mean "ten-sided polygon." This may be so because it has regularly been used that way over the last few years or decades, even if no one happens to use that word this week. The boundary of "the present time" is of course vague. There may be no precise point at which we can begin to say "people no longer do things that way." But when it is true, then it is no longer conventional to do things that way. It is no longer our practice.[14] The fact that "the present time" is vague entails that lexicographic decisions as to whether words have lost certain meanings will inevitably be uncertain, even when the evidence is complete.

Somewhat paradoxically, there may be a common way of doing something uncommon. How common a way W of doing A is, is determined by the *relative* frequency of W *given* A, not by the absolute frequency of W. For example, Americans seldom use the word "googol" or express the idea of 10^{100}. Nevertheless, using the word "googol" to express the idea of 10^{100} is both a common way of using that word and a common way of expressing that idea. Similarly, many words are considered profane, vulgar, obscene, offensive, or taboo. Their use is therefore unconventional. It is nevertheless conventional to use the abhorrent words to express certain ideas. There is a common way of expressing ideas it is improper to express. Similarly, the use of "ain't" is nonstandard. But when it is used, it is almost universally used to mean "am not." There is a common way of using words it is improper to use. "Archaic" words like "affright" are rarely used nowadays, and are considered old-fashioned. Nevertheless, when "affright" is used, it is commonly used to mean "frighten" or "fright." The fact that it is only the way in which we express ideas that is conventional, not their expression itself, holds even in the case of commonly expressed ideas. People commonly express the idea of food, for example, but this action is neither arbitrary nor self-perpetuating in the right way. The fact that many people before me have expressed the

14 Contrast Gilbert 1996: 73–4.

idea of food is a negligible reason for me to express it when I am hungry. The neo-Gricean analysis should be understood as requiring that *using e to mean* μ is a conventional way of both *using e* and *meaning* μ.

In some cases, such as driving on the right, a regularity will be socially useful provided that it is the *only* common way of doing things. But in other cases, such as language, the mutual interest may be served by several common ways of doing things. The general interest in communication is served quite well even though "plane" and "airplane" are both commonly used to express the idea of an airplane, and even though "plane" is commonly used to express both the idea of an airplane and that of a certain tool. It is consequently not the case that, by convention, people use "plane" *only when* they mean "airplane" or *whenever* they mean "airplane." As these examples illustrate, conventional regularities need not be perfect: a common way of doing things need not be universal or unique. It is conventional to drive on the right even though people occasionally drive on the left; and it is conventional to use "plane" to mean "airplane" even though people often use it to mean "tool used to plane wood." "Bank" is conventionally used to mean "row of keys on the organ," but it is much more commonly used to mean other things. The existence of ambiguity and synonymy is not problematic for the neo-Gricean analysis. Nor is the fact that there are many regional variations and specialized vocabularies in a language such as English. "Y'all" is commonly used to mean "you all" in English even though, southerners being in the minority, most speakers do not so use it. Even Lewis's (1969; 1975) final definition of a convention as something that *nearly everyone* does was much too strong. In Definition 9.1, "regularity" must denote patterns or commonalities rather than (near) universalities.[15]

The term "common" is of course vague, as are the near-synonyms "often" and "frequent." There is no precise percentage *r* such that a way of doing something is common in a population iff more than r percent of the population do it that way. We know that *r* must be sufficiently greater than 0, and that *r* may be considerably less than 100. A 30 percent incidence surely makes the way common, but it is hard to say whether 10 percent suffices. This vagueness is not a defect in Definition 9.1. For it is

15 Universality or near-universality is assumed in Arnauld 1662: 90; Grice 1968: 232–3; Schiffer 1972: 128–9, 136, 154, 156; 1987a: 250–1; Walker 1975: 169; Kempson 1975: 150–2; Bennett 1976: 213; Bach & Harnish 1979: 108–10, 189–95; Yu 1979: 282–5; Loar 1981: 256; Davies 1987: 717; S. Russell 1987: 730ff.; Avramides 1989: 68; and Chierchia & McConnell-Ginet 1990: 153. Contrast R. Harris 1980: 104–5; Gilbert 1996: Chapters 3–4.

just as hard to say when a self-perpetuating practice is common enough to count as conventional, or when a continuing usage is widespread enough to count as part of the language. If the students at Ridgeview High all begin using "nervy" to mean "intellectually gifted," that is not enough for it to be true that "nervy" means "intellectually gifted" in English. If the usage spreads from this high school to others, from high school to college, and from there to the popular media and so on, then eventually it will be true that "nervy" has come to mean "intellectually gifted" in English. But there is no point along this continuum that we can point to and say "Here is where the 'nervy' first meant 'intellectually gifted' in English." Consequently, lexicographers have to agonize over when a new usage is frequent enough to be included in their dictionaries, and when an old usage is dead enough to be labeled "obsolete" or dropped altogether. "Those dictionaries that have the most liberal policy of inserting new terms are precisely those most likely to be filled with antiquated slang and fad words of another generation" (Landau 1984: 162). The lexicographer's problem is compounded by the fact that the citation files on which such decisions are based constitute a very limited sample of the total domain of English usage (Landau 1984: 161–6, 220).

§9.6 CORRECT USAGE

We observed that conventional rules serve as generally accepted standards of correctness. Whether we are dealing with linguistic or nonlinguistic conventions, the members of a population in which an action is conventional generally believe the action to be right and proper. The criticism occasioned by violations of the norm and rewards gained by conformity are social pressures acting to conserve the convention. Embarrassment and pride play a similar role within the individual.

The normative force of linguistic conventions is complicated by the fact that a language such as English has many dialects, whose conventions differ to some extent. When different dialects coexist in the same region, residents often have to choose which conventions and which standards of correctness to adopt. Dialects typically differ in social status. Most languages have a prestige dialect, which is widely respected by all social levels because it is spoken and written by the most highly educated, powerful, and respected members of the society, especially on formal occasions and in official documents and serious literature. At the other extreme, the languages of the major colonial powers "lexicalized" a number of pidgins and Creoles, spoken by the lowest classes and regarded as corruptions of

the colonizing language. The conventions of the prestige dialect define *standard* usage. Speakers of the standard dialect exert pressure on others to make their speech conform to the standard conventions. Speakers of the other dialects are motivated to adopt standard usage by their desire to succeed and move up, to be powerful, well respected, and so on. For these same reasons, the standard dialect is taught in the schools. Editors revise manuscripts and speeches to conform to standard usage, and television and radio announcers are trained to suppress all traces of regionalism. Since lexicographers tend to rely predominantly on edited media for citations (Landau 1984: 164, 220) – and on their own erudite linguistic intuitions – dictionaries tend to preserve and reinforce standard usage even when trying to be "nonprescriptive."

Speakers to whom the prestige dialect is native tend to regard nonstandard usage as manifestations of ignorance or illiteracy. They often regard other dialects as inferior, degenerate, or corrupted. Many nonstandard speakers come to regard their own dialect in the same way, and seek to suppress their native habits. Given the arbitrariness of conventions, however, there is no more objective basis for judging nonstandard dialects to be intrinsically inferior than there was to judge English and other vernaculars to be inferior to Latin during the medieval and early modern periods. Some speakers of a nonstandard dialect learn to use the standard language in public discourse, and restrict their native dialect to private communication among friends and close colleagues. Indeed, in the working classes, pressure from close social networks to maintain the vernacular is greater than that from the upper classes to conform to the standard (Fasold 1990: 236–7). Linguistic conventions, like those of dress, serve as an instrument of *group identification*. Communication is not the only social interest served by linguistic conventions.[16]

It is particularly important to note that for e to mean μ in English it is not essential for e to mean μ in Standard English. That is, the use of e to mean μ need not be standard usage in any sense that favors one dialect over another. For example, "y'all" means "you all" in English even though its use is restricted to the South. More radically, "ain't" means "am not" in English even though its usage is nonstandard and widely considered to be incorrect (when it is not used to be deliberately colloquial or folksy). You and I will undoubtedly be reluctant to say that the use

16 Cf. Ullmann-Margalit 1977: §4.1 and Pateman 1982: 142. For a discussion of other interests served by linguistic conventions that play an important role in linguistic phenomena, see my *Implicature* (1998: Chapters 5–6).

of "ain't" is conventional, because that wrongly implies that we do not think there is anything wrong with it. But the use of "ain't" to mean "am not" is socially useful, self-perpetuating, arbitrary, and common in several regions,[17] thus fitting Definition 9.1. The usage is self-perpetuating by the force of precedent, association, and habit, and even by tradition and social pressure in areas where it is used. The perpetuation of this convention is impeded, however, by the force of the social sanctions imposed by the larger community, which presses to replace nonstandard usage with standard.

Occasionally, usage is considered to be incorrect even though it is perfectly conventional among speakers of the standard dialect.[18] My favorite examples are split infinitives and sentences ending with prepositions. Here is another, with some documentation.

> *Hopefully*, as used to mean it is to be hoped or let us hope, is still not accepted by a substantial number of authorities on grammar and usage. The following example of *hopefully* in this sense is acceptable to only 44 percent of the Usage Panel: *Hopefully, we shall complete our work in June.* (*American Heritage Dictionary* 1969: 634)
>
> *Webster's New World Dictionary (World)* says of *hopefully* (used to modify an entire clause): "regarded by some as a loose usage, but widely current." Granted that the scientific basis of such statements leaves much to be desired: we really do not know who or how many regard *hopefully* as a loose usage, apart from writers of usage books. . . . [But] when the editors of *World* say that *hopefully* is "widely current," they *know* it is widely current, at least in writing (and probably much more so in speech), because they have umpteen citations from all kinds of sources to prove it. (Landau 1984: 217)

People often do not recognize in an explicit manner what the actual conventions are. Indeed, people sometimes hear what they expect to hear,[19] and often would like to legislate their own linguistic preferences. In a

17 The *Oxford English Dictionary* describes "ain't" as belonging to Cockney, the dialect of those born in the East End of London; its use migrated to Appalachia, the inner cities, and elsewhere in America. According to the *Dictionary of American English on Historical Principles,* "ain't" is a contraction of "air not," the word "air" being a dialectal variant of "are" used as recently as 1884.

18 Buckley 1982 and Follett 1970 are noted prescriptivists. Contrast Hall 1962: 434–5; Read 1962: 438; Fries 1940; Smith et al. 1952: 274; Leonard 1929: Chapter 13; Labov 1970: 14–17; R. Harris 1980: 105–6; Bolinger 1980; 1982; Nunberg 1982; Landau 1984: 198; Love 1994.

19 Landau 1984: 202–3, following Labov 1966.

tradition that extends back to Quintilian (A.D. 35–95)[20] and that became firmly established in the eighteenth century, popular writers of usage manuals have erroneously held that correct usage is determined by etymology, analogy, logic, or the rules of classical languages (first Greek and then Latin). What is important for us is that meaning is determined by conventional usage even when accepted standards diverge from convention. "Hopefully" does mean "It is hoped" when it precedes a clause.

We have observed that the arbitrariness of conventions is sometimes obscured by their normativeness. In the case of language conventions, their arbitrariness is so salient that many linguists find their normative force hard to defend. Indeed, how can it be correct or proper to use words one way rather than another if the choice is arbitrary? How could there be any objective or rational basis for our judgments of correctness? These are important questions. To answer them, we must note that a *regularity in action* may be arbitrary in instances when a *particular choice of action* conforming to the regularity is not. It is the general use of "red" to mean "red" that is arbitrary. What makes this practice arbitrary is the existence of other possible regularities that would have been just as useful to society. If people regularly used "blue" to mean "red," our collective interest in communication would be served just as well. In contrast, *my* use of "red" to mean "red" is not at all arbitrary. The fact that I want to be easily understood together with the fact that my audience speaks English normally gives me a conclusive reason to use "red" rather than "blue" to mean "red."

Furthermore, judging speech to be correct or incorrect is like judging a musical performance to be correct or incorrect. A particular sequence of notes, say C-C-E-E-G-G-E, will be correct if I am trying to play Haydn's *Surprise* Symphony, incorrect if I am trying to play *Twinkle Twinkle, Little Star* (C-C-G-G-A-A-G), despite the similarity of the melodies. Similarly, using "rot" to mean "red" is correct if I am trying to speak German, incorrect if I am trying to speak English. Using "bonnet" to mean "roof of the car" is correct if I am speaking British English, incorrect if I am speaking American English. Using "red" to mean "blue" may even be correct if I am using a code that I have stipulated. In general, a musical performance is correct provided the musician performs the sequence of notes that defines the work he or she is trying to perform. A speech performance is correct provided the speaker uses "modes of expression"

20 Harris & Taylor 1997: 92–3.

defining the language he or she is trying to use (Chapter 7). That is, if S uses e to mean μ and is trying to speak L, then S's speech is correct if and only if e means μ in L. Conventional usage is the proper standard of correct English usage if the neo-Gricean analysis is correct, because conventional usage determines what words mean in English. That is, on the neo-Gricean analysis, conventional usage determines which modes of expression are in the system that constitutes English, and therefore determines whether the modes that S produced are among them. Linguistic correctness is thus as objective as it is relative. Correctness is in this respect like legality, although its source is convention rather than legislation.

The same point can be made using the notion of a rule. Judging the utterance of a sentence to be correct English is like judging a sequence of chess moves to be correct. The moves are correct if they are allowed by the rules of chess. The utterance is correct if the sentence is allowed by the rules of English. We can freely choose to speak one language rather than another, just as we can freely choose whether to play chess or checkers. Our choice determines which rules of correctness govern our actions.[21]

On the neo-Gricean analysis, the rules of English are determined by conventional usage. So conventional usage is the proper standard for judging whether or not speech is correct English. The linguistic conservatives who insist that conventional usage is incorrect are as wrong as the linguistic liberals who declare that native speakers cannot make mistakes.[22] Any notion of *absolute* correctness is unfounded, if that means that Latin is more correct than the vernacular, or British English more correct than American, or Standard American more correct than Southern. But the claim that A is using a language correctly does not entail the superiority of the language or dialect she is using over any other. It only claims that her usage is permitted by the language she is using. The question of whether a student should speak Black English ("Ebonics") or Standard English is not a question for linguistics, and cannot be settled by reference to the conventions of Black English, Standard English, or English. But *given* that a student is trying to use Standard English (or Black English), the question of whether a particular usage is correct or incorrect is determined by the conventions of Standard English (or Black English).

21 Cf. Alston 1964b: 141–2; 1974: 19; 29–30; 1994: 47.

22 This issue is framed in Follett 1970: 3 and Nunberg 1982: 36. Both views are to be contrasted with the appropriately nonjudgmental attitude expressed in Fasold 1990: 188–9. See also Knott 1934: 88, Labov 1970, and Bolinger 1980.

It should be noted that *correct* usage is not necessarily *good* usage, just as a correct performance of a piece of music is not necessarily a good performance. Correct usage may be bad writing if it is wordy, difficult to understand, poorly organized, too informal, excessively elliptical, obscene, or just ugly. It is much harder to avoid these defects and thereby write well than it is to write correctly. We might also debate whether good usage must be perfectly correct. But that issue would take us too far afield.

Correct usage, furthermore, is not necessarily *true* usage. If I say that President Bush is a bachelor, I am using "bachelor" correctly as long as I mean "unmarried male." But while my use of "bachelor" is correct English, my statement that Bush is a bachelor is incorrect, that is, false. We can use sentences correctly to express beliefs that happen to be false. We can also use sentences correctly to lie, joke, tell stories, or engage in hyperbole. The "problem of normativity" as discussed by Kripke (1982) in his exposition of Wittgenstein's views elides linguistic incorrectness and falsity.

> The problem of normativity is that we are disposed to make mistakes some of the time, and so we cannot simple-mindedly assume that some predicate F is true of a thing just in case we are disposed to predicate F of that thing. This problem arises at the level of the community as well as that of the individual; on occasion, the community is disposed to make mistakes *en bloc*. . . . (Lance & O'Leary-Hawthorne 1997: 312–13)[23]

Nothing in the neo-Gricean analysis implies that if everyone applies the term "entomologist" to an object, then that application cannot be mistaken. Two kinds of mistake are possible. First, as we noted in §8.5, it is quite possible for everyone to use F mistakenly to mean "etymologist" when in fact F means "entomologist." Second, even when correctly using "F" to mean "entomologist" it is possible for everyone to assert "N is F" falsely because N is not in fact an entomologist (or an etymologist, for that matter). Whether a term "F" applies to an object N depends on two factors: the meaning of "F" (or equivalently, the concept expressed by "F") and the nature of N. The conventional usage of "F" determines what "F" means, and therefore what "N is F" says about N. The facts about N then determine whether what "N is F" says about N is true or false. Once the meanings of its terms are fixed, conventional usage is irrelevant

23 See also Lewis 1969: 194–5; 1975: 51; Blackburn 1984: 821; Millikan 1984: 4–5, 52; Horwich 1998a: Chapters 8, 10; and §10.2 here.

to whether a sentence is true or false. If "F" expresses the concept of an entomologist, then "N is F" is true iff N is an entomologist.

§9.7 LEWIS'S EQUILIBRIUM CONDITION

Lewis's analysis was inspired by the fact that communication presents a "coordination problem." Speakers must use a language that their audience understands if communication is to occur. Any language will do, as long as they share it. The use of the same language by both speaker and hearer thus resembles an "equilibrium solution" in the game theoretic sense: no one can improve his or her outcome by unilaterally acting differently. Lewis accordingly held that a regularity R is conventional only if "[t]here is a general preference for general conformity to R rather than slightly-less-than-general conformity – in particular, rather than conformity by all but any one" (1975: 5).[24] This condition fits the driving case very well: given that nearly everyone drives on the right, we have a very strong preference that everyone do that.

Lewis's equilibrium condition is much too strong for most conventions. It does not fit conventions of dress, for example. We generally do not have to coordinate our clothing to achieve any ulterior end. Organized sports (e.g., baseball uniforms) and certain professions (e.g., police uniforms) are notable exceptions. It is conventional for women to wear dresses, even though it is far from universally true these days that almost everyone has a strong preference that all women wear dresses. The degree of conformism that characterized the 1950s represents a cultural extreme. As S. Miller (1986: 122) observes, dress practices would not cease to be conventional should most people want a few individuals to make complete fools of themselves by dressing inappropriately.

The equilibrium condition also fails for linguistic conventions, as the examples of ambiguity given earlier show. The condition does not even fit the use of a particular language. The English language community is so large, for example, that we really do not much care whether absolutely every member speaks English, given that almost every member does. The only people we really care about are those we wish to communicate with. And we will have a strong preference that all of them speak English only if English is the only language we know. Since a substantial percentage

24 See also Lewis 1969: 78. Cf. Schelling 1960: 91–106; Ulmann-Margalit 1977: Chapter 3; Gilbert 1996: Chapter 3. S. Miller (1986: 131) rejects Lewis's condition but replaces it with another that is similarly objectionable. Contrast Jamieson 1975: 77 and Nunberg 1981: 219.

of English speakers are bilingual, it is not true that almost all have the conditional preference characterizing the equilibrium condition.

§9.8 MUTUAL KNOWLEDGE

Lewis and his followers argued that for an action to be conventional, it must be *mutual knowledge* that the action is regular, socially useful, self-perpetuating, and arbitrary.[25] Some features of a conventional action are undeniably "common knowledge" in the ordinary sense of this term. It is generally known, for example, that people use the word "red" to mean "red." But Lewis and his followers gave "mutual knowledge" a highly technical sense. Schiffer's definition is representative: it is mutual knowledge that p iff everyone knows that p, everyone knows that everyone knows that p, everyone knows that everyone knows that everyone knows that p, and so on. Given that mutual knowledge of one proposition requires knowledge of an infinite sequence of ever more complex propositions, there has been considerable debate as to whether mutual knowledge is logically or humanly possible. I think not, but we need not quibble about such metaphysical details. If we are concerned with the conventional use of words, the mutual knowledge condition must be rejected for perfectly elementary reasons: *linguistic conventions need not be universally known*. In that event, even the *first-order* condition of mutual knowledge is unsatisfied. The problems cannot be eliminated by weakening mutual knowledge and reducing it to mutual belief. The problem is lack of belief rather than lack of truth or justification.

For example, satisfaction of the arbitrariness clause may fail to be mutual knowledge.[26] Speakers of English need not realize, for example, that word order is arbitrary. They may mistakenly think that English word order is "natural" – as Arnauld and Diderot thought about French (Chomsky 1965: 6–7). Lyons (1971: 4) reports that ancient Greek grammarians actually believed that there was a necessary connection between the meaning

25 Lewis 1969: 52–68, 152–9; 1975: 6; 1976; Schiffer 1972: 30, 131, 148–55; 1987a: 249–61; Bennett 1976: 179–80, 190–2; Loar 1976a: 353; Cummins 1979, 352; Bach & Harnish 1979: 108–10, 189–195; Hungerland & Vick 1981: 50, 137–44; Pateman 1982; S. Russell 1987: 730–1; Hirschberg 1991: 11; S. Miller 1992: 438; and Laurence 1996: 277. Antecedents include Hume 1739: §3.2.2 and Schelling 1960: 54, 89–94. Contrast Jamieson 1975; Blakemore 1987a: 28, McCawley 1987: 723; and Gilbert 1996: Chapter 1. Gilbert (1996: 110, 198) claims it must be common knowledge that people believe they ought to perform the action.

26 This has been observed in Burge 1975; Jamieson 1975: 78–9; Grandy 1977; and Blackburn 1984: 121.

of a word and its form. To take a more contemporary example, the mere fact that Chomsky and others *believe* that certain linguistic universals are innate does not suffice to prove that they are not conventional.

Even more elementary – and decisive – is the fact that satisfaction of the regularity clause need not be mutual knowledge. Not everyone knows, for example, that people use the word "googol" to mean 10^{100}. Indeed, the percentage of English speakers who know anything about the word "googol" is quite small. Since few people know what "googol" means, *no one* knows that *everyone* knows that people use "googol" to mean 10^{100}, no one knows that everyone knows that everyone knows this, and so on. No one even knows that *most* people know it. All of this holds for "implicit" or "potential" knowledge as well as for actual, explicit knowledge. There is nothing special about the word "googol" here. No one knows every word of a natural language like English, and relatively few words are known by all or nearly all speakers of English. Those who have a large vocabulary are usually intelligent enough to realize that most people do not.

The second-order knowledge condition is also problematic. Presumably, the group with mutual knowledge is the entire group of English-speaking people. Since this group has around a billion members of all ages spread all over the globe, I am not sure there is anything short of the most trivial tautologies that I believe all or nearly all of its members know.

I believe that the requirement of mutual knowledge was built into Lewis's definition of convention for three reasons. First, Lewis needed the mutual knowledge condition to prove, as a game-theoretic theorem, that ideally rational agents will choose the conventional solution to coordination problems. But no one is ideally rational. No one has complete knowledge of, or reasons flawlessly in, every situation requiring a decision. Moreover, many highly rational people fail to choose the conventional solution, or choose it either for nonrational reasons or by chance. We do not always assume that others are rational, much less as clever as we are. No one, moreover, can deduce that other intelligent agents will choose a particularly salient solution of a coordination problem without making a highly defeasible assumption about how the others will respond to saliency.[27] It is more faithful to the conventional meaning of "conventional," I believe, to drop the requirement of mutual knowledge from the definition, and to reformulate Lewis's theorem conditionally. What Lewis has shown is that ideally rational agents with

27 Cf. S. Miller 1992: 436; Gilbert 1996: Chapters 1, 2, 4.

mutual knowledge will choose the conventional solution to coordination problems.

Second, Lewis needed the mutual knowledge condition to prove that if S followed a signaling convention in using e to perform certain actions, then S satisfied the Gricean definition of speaker meaning. Since there is abundant reason to reject the Gricean definition, this theorem also has limited interest.

Third, and perhaps most important, the requirement of mutual knowledge was primarily motivated by the desire to exclude recherché cases reminiscent of those inspiring the Gricean higher-order intention clauses (see §4.3).[28] The cases are purely hypothetical, and I do not believe that they are clear enough to warrant discussion. We can say with complete confidence, though, that if they are not cases of convention, it is not because of the lack of mutual knowledge. For that is rarely, if ever, present in the clearest cases of conventions.

28 Lewis 1969: 59; Grandy 1977; Blackburn 1984: 121.

10

Compositional Word Meaning

According to the neo-Gricean analysis presented in Chapter 8, word meaning is conventional cogitative speaker meaning. That is, what an expression means is determined by what idea people conventionally use it to directly express. This thesis works very well for individual words, stock phrases, dead metaphors, and idioms. "Green" means green only because it is conventional for English speakers to use the word "green" to mean the color green. The phrase "kicked the bucket" means "died" only because people conventionally use it to mean that. The basic neo-Gricean analysis does not, however, account for the *constructive* and *recursive* character of the semantic rules of a language.[1] That is, the analysis does not account for the way in which the meaning of a complex word, phrase, clause, or sentence is normally determined by the meanings of its components. In this chapter we will account for the compositionality of linguistic meaning by adding a recursive element to the neo-Gricean analysis. This element is provided by the fact that there are conventions whereby word structures are used to express idea structures. We will address objections that have been leveled against the notion of compositionality, and argue that projection rules need not be restricted to block generation of "anomalous" meanings.

1 Cf. Geach 1957a: 12; Fodor & Katz 1964: 11; Ziff 1960: 59–62; 1964: 391; 1967: 6; Davidson 1967: 79; Rosenberg 1974: 3, 30; Chomsky 1975: 74–5; Grandy 1977: 135–6; Searle 1979: 156; Platts 1979: 89; Cummins 1979: 350; Harrison 1980: 190–6; Loar 1981: 256; Blackburn 1984: 10–18, 35–6, 127–33; Hornstein 1984: 123; Schiffer 1987a: Chapters 7–8; Avramides 1989: 4–13; Chierchia & McConnell 1990: 6–7, 152; Katz 1987: 171; Grandy 1990; Horwich 1998a: §3.9. Contrast Fodor 1998a: 9.

The deficiencies of the neo-Gricean analysis can be seen most starkly by reflecting on the fact that every natural language contains a large – indeed, infinite – number of meaningful sentences that have never in fact been uttered. Some are too long or convoluted ever to be uttered. Furthermore, whereas the finite stock of words in a language is generally used over and over again, it is unusual for the very same sentence to be used twice. If a meaningful sentence has been used no more than once, however, then it is not conventional for people to use it for anything (§9.5). So the neo-Gricean analysis wrongly entails that the sentence has no meaning.

Even when we confine our attention to sentences that are used repeatedly, these may have meanings that people seldom, if ever, intend. Consider (1):

(1) He gave him hell.

It is conventional to use (1) to mean "He gave him a tongue-lashing," not "He gave him the place where sinners go." (Zeus gave Hades to his brother. But (1) is not customarily used even for that transaction.) So the neo-Gricean analysis rules that (1) means the former but not the latter. Yet (1) has both meanings in English.[2] Katz and Fodor (1963: 496–8) assumed that if a sentence is not understood in a "possible" sense, then the sentence does not actually have that sense. But even though (1) is not normally understood as meaning "He gave him the sinners' place," it *can* be understood that way. And sometimes (1) *must* be interpreted that way, as in "God's brother Theo gets everything he asks for; Theo asked God for the place where sinners go; therefore, He gave him hell." In fact, once we engage in reflections such as this, it is hard to *stop* understanding (1) in the way that Katz and Fodor would dismiss. While such reflection changes the way we "hear" a sentence of English, it does not change the English language. It did not just create a new sense for (1).[3]

2 Grice's (1968: 127) strategy suffers from the same two defects. Neale (1992: 553–4) faulted Grice's inability to handle sentences with indexicals (a problem that can be solved by switching from cognitive to cogitative speaker meaning – §8.5), but evidently did not see the problems that we are currently discussing.

3 Stillings et al. 1995: 373. In an interesting departure from their general Chomskyan framework, Katz and Fodor thought that the output of a semantic theory should match the *performance* of a fluent speaker, rather than his *competence*. Something like Katz's and Fodor's "selection rules" will certainly be necessary for a *cognitive* theory of understanding, however. See §10.6 and §10.8.

Productivity is of primary importance, but it should not be overblown. Ziff (1967) claimed, for example, that Grice's (1957) analysis "lacks the basic ingredients of meaning: a set of projective devices." As Patton and Stampe (1969) observed, Ziff confused an important feature of natural languages with an essential feature of meaning. Blackburn (1984: 36) seems to have made the same mistake. Signals may have meaning even though they are part of no projective system, and unstructured words and idioms have meanings that do not result from projection. The problem is not that all word meaning is projective, but that some is. That case, moreover, is central to the functioning of all natural languages and many artificial languages.

§10.2 THE POTENTIAL MEANING ANALYSIS

One solution to the problem of unused expressions is to equate their meaning with what people *could* mean by them, or *would, if* they used them to mean anything.[4] With "would," the suggestion is too strong: it would rule, for example, that (1) does not mean "He gave him hell (the place)," since no actual speaker would use it to mean that. It would also entail that "He went to a bank" does not mean anything, since it is not true that people *would* mean "He went to a commercial bank," nor that they would mean "He went to a river bank," and so on for all of the meanings of that sentence. In short, the proposal does not allow for the possibility of ambiguity. Finally, sentences may be too complex for a speaker to mean what they mean.

With either "would" or "could," the suggestion is much too weak. People could mean anything by "Snow is white," since they could use it as part of a code. And even though "The table who annoyed me here found" cannot be said to mean anything in English, a struggling speaker of English could use it to mean "The table that annoyed me was found here." Similarly, anyone who said "He singded a little good" would probably mean "He sang pretty well." But "singded" is not an English word, and "little" cannot modify "good" in English. Finally, even though "He's a son of a stickleback fish" means that he is the male offspring of a fish of

4 Cf. Chomsky 1961a; 1961b; Ziff 1964; 1967; Katz 1964a; Stampe 1968: 171; Patton & Stampe 1969: 13–15; Schiffer 1972: 7, 119, 156; 1982: 122–3; 1987a: 214–15; Platts 1979: 89–90; Biro 1979: 242; Loar 1981: 254–5; Hungerland & Vick 1981: 69; and McDermott 1988: 36. Somewhat stronger variants are presented in Katz 1977b: 14–15 and Bach 1987a: 70. Many problems cited in this literature can be avoided by focusing on what people would mean *cogitatively*, such as Ziff's "son of a stickleback fish" problem, and Schiffer's "grr."

the species *Eucalia inconstans*, it could be the case nevertheless that people *would* use it to mean only "He's an SOB."

Even when this counterfactual neo-Gricean analysis assigns the proper meaning, it does so *per accidens*. I presume that no one (other than me, here) will ever use (2), simply because the appropriate occasion will never arise.

(2) Bill ate purple spaghetti with a titanium tea cup.

I am confident, nevertheless, that if people did use (2), they would use it to mean "Bill ate purple spaghetti with a titanium tea cup." So the current proposal assigns the proper meaning to (2) in the actual case. But let us imagine a case that is counterfactual in one respect: people would not use (2) unless they first constructed and then used a code in which (2) means "Snow is white." The proposal rules that (2) means "Snow is white" in the imagined case. This is incorrect, since, for one thing, I implied that the code in question had not actually been constructed. The underlying problem is that the proposal focuses on a consequence of word meaning, rather than on its essence. People would use (2) to mean "Bill ate purple spaghetti with a titanium tea cup" because (2) has that meaning in English, because people know English, and because people have a certain range of motivations. A difference in the motivational factors alone would alter people's dispositions without changing what (2) means.

We might try to improve these proposals by equating the meaning of an expression with what people could mean by it *if they were following the conventions of the language*.[5] But what are the relevant conventions for a sentence like (2), which itself has no conventional use? The most obvious candidates are the conventions governing the individual words making it up, and the conventions governing their mode of combination. Since people could mean "He gave him hell (the place)" by (1) without violating the conventions governing its components and their mode of combination, it follows that (1) has the meaning that it does. Since people would mean "Bill ate purple spaghetti with a titanium tea cup" by (2) if they conformed to the conventions governing its components and their mode

5 Cf. Alston 1971: 35–6; 1974: 35–40; 1977: 17, 21, 23; 1980: 124, 127–8; 1987: 151, 157–8; 1994: 29, 36–8; Schiffer 1972: 156; 1987a: 12, 250–1; Walker 1975: 168–9; and Blackburn 1984: 127–33. Schiffer's formulations, however, rule out ambiguity. Lance and O'Leary-Hawthorne raise the "Kripkensteinian" objection that "[g]iven any finite fragment, there will be a number of sets of rules that deliver that fragment but which yield radically different results for sentences outside the fragment" (1997: 293). Only some of those rules will count as conventions.

of combination, it follows that (2) has the meaning that it does. So this solution seems to work.

Chomsky (1975: 74) claimed baldly that "there are no practices, customs, or habits . . . that carry us very far in accounting for the normal creative use of language." Griceans have had a problem refuting this denial, given their fixation on cognitive speaker meaning, and their consequent acceptance of the primacy of sentence meaning (§8.4). As a result, Griceans could not say that the relevant conventions governing sentence parts were the conventions to use those parts to mean certain things. Having shifted from cognitive to cogitative speaker meaning, we can. We can say, ever so naturally, that (2) means what it does because people conventionally use "cup" to mean "cup," "tea" to mean "tea," and so on. Since we have defined speaker meaning in terms of expression, we can add that certain grammatical constructions are conventionally used to express certain conceptual structures. We do not, therefore, have to resort to the hand waving characteristic of Griceans at this point. As an added bonus, we can simplify the definition. Since we are able to specify the relevant conventions, it is superfluous to refer to what people would or could mean when following them. We can define word meaning directly in terms of the conventions.

§10.3 THE RECURSIVE NEO-GRICEAN ANALYSIS

We will use the definiens of the neo-Gricean analysis as the *base clause* of a recursive definition. This clause gives the meaning of the smallest meaningful units in the language, as well as the idiomatic meaning of pure idioms, dead metaphors, and other noncompositional word combinations.[6] Thus "red" means "red" because speakers conventionally use "red" to mean "red." "France" means "France" because it is the customary practice to use "France" to mean "France." "Kicked the bucket" means "died" because that is what speakers conventionally use it to mean. An analogue of the neo-Gricean analysis can be used to define the meaning of the basic grammatical constructions used in the language, which provides the *recursion clause*. To vary an earlier example, a phrase of the form "N V-er" means "person who Vs Ns" rather than "V-er who is an N" because it is conventional for speakers to use such phrases to mean the former rather than the latter. We can now say that the conventions

6 Cruse (1986: Chapter 2) and Cowie (1992) survey the broad range of "ready-made memorized combinations."

Table 10.1. *Two expression structures*

N	V-er	E_1[N,V-er]	E_2[N,V-er]
cat	*love*	*cat lover*	*lover of cats*
dog	*hate*	*dog hater*	*hater of dogs*
violin	*play*	*violin player*	*player of violins*
etc.			

of English assign meanings directly to "girl" and "watcher" and to the construction "N V-er." These conventions indirectly assign the meaning "Person who watches girls" to "girl watcher." In the terminology of §7.5, the base clause of the definition will be satisfied in virtue of the *lexical* rules of the language, while the recursion clause will be satisfied in virtue of the *construction* rules.[7]

We used "N V-er" and "something that Vs Ns" to represent certain *expression structures*. We also used them indirectly to represent certain *idea structures*. For formal purposes, it is more convenient to use functions to represent such structures. A familiar example is the propositional function of quantification theory, which in some treatments represents a sentence form that results in a sentence when names or individual constants are substituted for individual variables. Thus the sentence form "x is a composer," whose substitution instances are "Brahms is a composer," "Chopin is a composer," "Dvorak is a composer," and so on, can be represented by the function C(x) – or Cx, for short – whose domain is the set of individual constants or names and whose range is the set of sentences combining them with the predicate "is a composer." In general, let $E[x_1, x_2, \ldots, x_n]$ designate a function from a number of expressions to a complex expression that contains them. The form "N V-er" corresponds to the function E_1[N, V-er] (see Table 10.1). The related form "V-er of Ns" is represented by the function E_2[N, V-er], whose range

7 Cf. Ziff 1960: 61–2; Katz & Fodor 1963: 482, 493, 503–16; Katz 1964b: 520; 1972; Schiffer 1972: 161–2; Evans 1973: 303; Blackburn 1984: 133; Horwich 1998a: Chapter 7. Schiffer (1987a: 251) has proclaimed "the hopelessness of stating for each primitive vocabulary item and syntactical construction of the language a separate practice that will yield a finite set of practices that together will determine the meaning of each of the infinitely many sentences of the language. The difficulty will emerge in the need to refer without circularity to each of the other practices in stating the practice for any given word or construction." Unfortunately, Schiffer provided no evidence for this claim. Our ability to provide countless examples such as those given in the text seems to belie Schiffer's skepticism. I suspect that Schiffer overlooked the fact that conventional regularities need not be universal, exceptionless rules. He may also have been misled by then-current views of anomalous sentences (§10.8).

Table 10.2. *An idea structure*

i(N)	i(V-er)	I[i(N), i(V-er)]
i(*cat*)	i(*lover*)	i(*person who loves cats*)
i(*dog*)	i(*hater*)	i(*person who hates dogs*)
i(*violin*)	i(*player*)	i(*person who plays the violin*)
etc.		

contains "lover of cats," "hater of dogs," "player of violins," and so on. As the "of" indicates in this simple example, there is no requirement that $e_1, e_2, \ldots, e_n$ be the only components of $E[e_1, e_2, \ldots, e_n]$.

Let $I[i_1, i_2, \ldots, i_n]$ designate a function from a number of ideas to a complex idea containing $i_1, i_2, \ldots, i_n$. We will use $i(e_i)$ to designate the idea that the speaker directly expressed by e_i when using $E[e_1, e_2, \ldots, e_n]$. Thus it is conventional for English speakers to use both E_1[N, V-er] and E_2[N, V-er] to express I[i(N), i(V-er)], as specified in Table 10.2. We can use these functions to provide a recursive definition for word expression.

(3) Expression e expresses i iff (i) people conventionally use e to directly express i, or (ii) people conventionally use $E[x_1, x_2, \ldots, x_n]$ to directly express $I[i(x_1), i(x_2), \ldots, i(x_n)]$, where $e = E[e_1, e_2, \ldots, e_n]$, $i = I[i(e_1), i(e_2), \ldots, i(e_n)]$, and $i(e_i)$ is the idea expressed by e_i for all i from 1 to n.

Clause (ii) tells us that "aardvark lover" expresses the idea "person who loves aardvarks" because people conventionally use E[N, V-er] to directly express I[i(N), i(V-er)], use "aardvark" to directly express the idea "aardvark," and use "lover" to directly express the idea "lover." Note that (3) does not require, à la Grice (1968: 235–6), the existence of a convention to use $E[e_1, e_2, \ldots, e_n]$ that is in any way a "resultant" of the conventions governing $e_1, e_2, \ldots, e_n$. Examples like (2) show that there will often be *no* convention to use the complex expression.

Commonly used compounds like "cat lover" satisfy clause (i) as well as clause (ii). But expressions like "aardvark lover," which are rarely if ever used, will satisfy clause (ii) but not (i). The dead metaphor "left no stone unturned" gets its metaphorical meaning, "tried everything," from clause (i), and its literal meaning, "turned every stone," from clause (ii). "Kicked the bucket" similarly gets its idiomatic meaning, "died," from clause (i) and its regular meaning, "struck the bucket with a foot," from (ii). If $E[e_1, e_2]$ represents the structure of "kicked the bucket," then there is an idea structure $I[i_1, i_2]$ such that it is conventional for speakers to use expressions of the form $E[e_1, e_2]$ to express $I[i(e_1), i(e_2)]$, where I[i(*kicked*),

i(*the bucket*)] is the idea of striking with a foot the contextually indicated bucket. In contrast to "kicked the bucket," "cooked one's goose" is not a pure idiom. For it is a regular instance of the construction "cooked N's goose." "Cooked Aristotle's goose" cannot get its meaning from clause (i), because it has probably never been used. But expressions of the form "cooked N's goose" are conventionally used to express the idea structure "ruined N's chances." So "cooked one's goose" and "cooked Aristotle's goose" get their meaning from clause (ii).

With complex expressions like (2), we can work our way up recursively from the simplest components to the whole compound. For example, "spaghetti" is a commonly used word, so it is assigned a meaning by (i). Expressions of the form "purple N" are commonly used to express ideas of a certain structure. Given the meaning already assigned to "spaghetti," clause (ii) assigns a meaning to "purple spaghetti." Expressions of the form "ate NP" are commonly used to express ideas of a certain structure. Given the meaning already assigned to "purple spaghetti," clause (ii) assigns a meaning to "ate purple spaghetti"; and so on. Formulation (3) thus shows how, and to what extent, the meaning of a sentence, or other grammatical compound, depends on the meanings of the words making it up.

As the simple example we have been using illustrates, there is no requirement that the structure of $I[i(e_1), i(e_2), \ldots, i(e_n)]$ be isomorphic to the structure of $E[e_1, e_2, \ldots, e_n]$.[8] The structure of the phrase "aardvark lover" is its *surface structure*. The structure of the idea "person who loves aardvarks" could naturally be termed the *deep structure* associated with "aardvark lover." Three common linguistic phenomena prove that the structure of the idea expressed by a compound expression may diverge from the structure of the expression. First, complex phrases with different surface structures may express the same idea, as "aardvark lover," "lover of aardvarks," and "person who loves aardvarks" illustrate. Second, expressions with no surface structure may express structured ideas. Thus "vixen" expresses the idea of a female fox, and statement constants in propositional logic are used to express complete thoughts of unlimited complexity. Third, complex phrases are often amphibolous: ambiguous

8 Contrast Horwich, whose theory of compositional meaning requires that the meaning of a complex expression have the same structure as the expression: they must both result from applying the same combinatorial procedure P, to meanings (concepts) in the one case and to expressions in the other (see 1998a: 180). His theory also rules that only complex expressions can express complex concepts (1998a: 163–4), which means that "bachelor" could not possibly be synonymous with "unmarried man," and that abbreviations, stipulations, and codes are impossible.

in ways that do not result simply from ambiguities in their components. Thus "Russian teacher" can mean "teacher of Russian" or "teacher who is Russian." And "flying planes" can mean "the act of flying planes" or "planes that are flying."

Some expression structures do not have a conventional usage even though more specific structures do. Consider, for example, the general structure "ADJ N," consisting of an adjective followed by a noun. Is it conventional to use expressions of this form to express the idea "N that is ADJ"? That depends on whether the adjective is "regular." We recognize that when adjectives like "red," "large," "good," "fat," "smoky," "intelligent," and so on are combined with nouns, the result is conventionally used to express the idea "N that is ADJ." But when adjectives like "phony," "former," "would-be," "nominal," and so on are used, the combination is not conventionally used to express the idea "N that is ADJ." Whereas an old Vermeer is a Vermeer that is old, a phony Vermeer is not a Vermeer, and so cannot be a Vermeer that is phony. The conventions governing the two adjective categories ADJ_r and ADJ_i are different. Expressions of the form "ADJ_r N" are conventionally used to express ideas of the form "N that is ADJ_r." There is no general convention governing the use of expressions of the form "ADJ_i N." Instead, the conventions vary with the adjective. "Phony N" is conventionally used to express the idea "object that purports to be N but is not"; "former N" is conventionally used to express the idea "object that used to be N but is no longer"; "would-be N" is conventionally used to express the idea "person who hopes to be N"; and so on. The adjective "apparent" is ambiguous, with a regular meaning ("evident") and an irregular meaning. On the latter, "apparent N" is used to express the idea "object that appears to be N." The expression "Russian teacher" is ambiguous, we can now observe, because it can be used as an instance of "N V-er" or as an instance of "ADJ_r N." Hence clause (ii) and the two illustrative construction rules just discussed assign it both the idea of a teacher of Russian and the idea of a teacher who is Russian.

Formulation (3) explicitly defines word expression in terms of speaker expression. Definition 7.4 has already defined word meaning in terms of word expression: *e means* μ_i *iff e expresses i*, where "μ_i" stands for expressions expressing i. Alternatively, we can derive from (3) a principle defining word meaning directly in terms of conventional speaker meaning. Let us stipulate that $E'[x_1, x_2, \ldots, x_n]$ expresses $I'[i(x_1), i(x_2), \ldots, i(x_n)]$ in the metalanguage for every n-tuple of expressions $x_1, x_2, \ldots, x_n$. And let us say that speakers use $E'[x_1, x_2, \ldots, x_n]$ to mean $E'[x_1, x_2, \ldots, x_n]$ just

in case they use expressions of the form represented by $E[x_1, x_2, \ldots, x_n]$ to directly express the corresponding idea $I'[i(x_1), i(x_2), \ldots, i(x_n)]$. Then (3) is equivalent to (4) given Definition 7.4.

(4) Expression e means μ_i iff (i) people conventionally use e to mean μ_i, or (ii) people conventionally use $E[x_1, x_2, \ldots, x_n]$ to mean $E'[x_1, x_2, \ldots, x_n]$, where $e = E[e_1, e_2, \ldots, e_n]$ and the idea μ_i = the idea $E'[e_1, e_2, \ldots, e_n]$.

We will refer to (3) and (4), along with a minor revision made in §10.7, as the *recursive neo-Gricean analysis.* For short, the analysis may be characterized as saying that word meaning (expression) is *constructive conventional cogitative speaker meaning (expression).*

Given the use-interpretation, expression-communication, and expression-indication equivalences (Postulates 8.1–8.4), the base and recursion clauses of the neo-Gricean analysis can be reformulated in terms of interpretation, production, communication, and indication. Expression e means μ by the base clause if people conventionally use e to mean μ; or if people conventionally use e to produce or communicate the idea μ; or if people conventionally interpret others to mean μ by e; or if the use of e is a conventional and direct indication of idea μ. Moreover, if any of these conventions exist, all of them exist. Similarly, people conventionally use $E[x_1, x_2, \ldots, x_n]$ to mean $E'[x_1, x_2, \ldots, x_n]$ iff people conventionally understand expressions of the form $E[x_1, x_2, \ldots, x]$ as meaning $E'[x_1, x_2, \ldots, x_n]$. These basic conventions are connected in this way because the common interest sustaining such conventions is communication, and because the communication of ideas implies their transmission.

The recursive neo-Gricean analysis allows it to be true in general that the meaning of an individual word is its contribution to the meanings of sentences containing it. The analysis also allows it to be true in general that the meaning of a sentence is the product of the meanings of the individual words making it up and the way they are put together. But the recursive neo-Gricean analysis allows exceptions to both rules, and uses neither to define linguistic meaning. That is, it affirms neither the primacy of sentence meaning nor the primacy of word meaning. We saw in §8.4 that sentential primacy is incompatible with the general determination of sentence meaning by component word meaning. The fact that idiomatic sentences like "Hello!" "Good bye!" and "God damn it!" are exceptions to compositionality shows that subsentential word meaning is not primary either. Idiomatic sentences get their meaning directly by the base clause (i) rather than by the recursion clause (ii). Instead of taking either sentence

meaning or individual word meaning to be primary, the recursive neo-Gricean analysis defines both independently in terms of cogitative speaker meaning or the direct expression of ideas.

I have said that there must be conventional construction rules, which assign meanings to complex expressions on the basis of the meanings of their components. Such rules are necessary to account for the meanings of novel word complexes, and for unusual meanings of common complexes, on the basis of conventional speaker meaning. I have not said, however, that a semantic theory must use such rules. The most that follows from (3) or (4) is that however a semantic theory assigns meanings to expressions in the language, the assignments must conform to some set of conventional construction rules. For example, it is conventional for speakers of English to use expressions of the form "V-ing N" to express either the idea of an N that is V-ing, or the idea of the act of V-ing N. Because of these construction rules, the sentence "Flying planes can be dangerous" has its well-known ambiguity. That ambiguity can be accounted for within a generative semantic theory by the fact that the sentence can be derived from two different deep structures à la Chomsky.

I have also not said that the construction rules must be *discovered* by any form of "simple induction" from observed linguistic regularities. Indeed, nothing I have said implies that the construction rules have to be discovered in any particular way, or at all. Finally, I have not said (or denied) that the conventional constructions rules must be used in any by way the speaker or hearer, or that cognitive theories of language production and comprehension must attribute knowledge of them to the speaker or hearer.[9]

Loar seems to have anticipated my proposal in the closing pages of his *Mind and Meaning*, only to reject it in favor of the apparently speculative

9 According to Schiffer (1987a: 192), "it involves only a little exaggeration to say that among those who hold the CS theory for a considered reason, that considered reason is . . . the conviction that without the assumption of a compositional semantics there could be no explaining a person's ability to comprehend indefinitely many utterances of indefinitely many novel sentences." Compare and contrast Grandy 1990: 558–9. As I see it, the existence of a compositional semantics *is* implied by the ability to *correctly* understand indefinitely many novel sentences. But that is only because S correctly understands σ only if S takes σ to have one of the meanings it actually does have. So the ability to correctly understand novel sentences entails that novel sentences have meanings. And the best account of what it is for a novel sentence to have a meaning postulates a compositional semantics. The Davidson–Dummett attempt to avoid meaning by talking about understanding was, I believe, a failure. Cf. Schiffer 1987a: 205–6.

hypothesis that an unused expression e means μ provided that the generative grammar "entrenched" in speakers assigns that meaning to e.

> Within a communication intention theory, a grammar maps syntactic structures and words, roughly, onto structures and constituents of the beliefs sentences express. . . . Now what might make such a given correlation of structures and constituents *entrenched* in the usage of a population? . . . The ground level is patterns of usage, that is, patterns in the structure of utterances and of accompanying intentions; the upper story is a psychological internalization of those patterns, and its effect on their continuation in the usage of P. . . . Now the ordinary commonsense concept of *knowledge* of the rules of a grammar is not the right idea here. But we do have the idea, at least in a general way, of how a cognitive psychology might contain a generative procedure acquired as a result of the impact of usage, and which plays an appropriate role in the continuation of that usage – that is, in those correlations of syntactic structures and the functional structures of communication intentions. This would vindicate the intuitive idea that *conventional* regularities are at work in language; for the pattern – usage, recognition of usage, and impact of recognition on continued usage – would then correspond to something objective. . . . It seems likely that if the Chomskyan psycholinguistic picture turned out not to be vindicated, literal meaning would have to be counted as an arbitrary extrapolation from structural correlations in usage, one that arises from our simplifying proclivities as theorists. (Loar 1981: 257–9)

Loar does not say more to clarify his view, so I am not sure that I fully understand it. But let me make the following comments. As theories designed to generate assignments of meanings to sentences, grammars may "contain" all sorts of rules in a variety of ways. A grammar could, for example, state the mathematical or logical rules used in the process of generating the assignments of meanings to sentences. Not all of the rules used in a grammar, therefore, need be entrenched in the speaker of the language. Which ones do?

I think the only rules that need to be entrenched, for the purposes of reducing word meaning to speaker meaning, are the conventional lexical and construction rules. That is, whatever "entrenchment" is exactly, it must satisfy the following constraints: any rule (or combination of rules) assigning expression e to idea i is entrenched if speakers conventionally use e to directly express i, and any rule assigning expression structure $E[x_1, x_2, \ldots, x_n]$ to idea structure $I[i(x_1), i(x_2), \ldots, i(x_n)]$ is entrenched if speakers conventionally use expressions of the form $E[x_1, x_2, \ldots, x_n]$ to directly express ideas of the form $I[i(x_1), i(x_2), \ldots, i(x_n)]$. That is, conventional usage must be a sufficient condition of entrenchment. If some of the conventional lexical rules are not entrenched, Loar's theory will

fail to assign the proper meanings to morphemes like "red" or to idioms like "kicked the bucket." If some of the conventional construction rules are not entrenched, Loar's theory will fail to assign the proper meanings to grammatical compounds like "Bill ate purple spaghetti with a titanium tea cup." If any lexical or construction rules are entrenched *other* than those that are conventional, it will be at best a coincidence if Loar's theory assigns the proper meanings. If there is to be any guarantee that Loar's theory assigns the proper meanings, then conventional usage must be a necessary as well as a sufficient condition of entrenchment. In that event, Loar's theory will be equivalent to (3). Loar's theory will essentially amount to the completely nonspeculative claim that the members of a language community must have the psychological constitution necessary for their conventional linguistic behavior to be conventional.

It needs to be emphasized that when we say (2) means what it does in English because "spaghetti" is conventionally used to mean "spaghetti," because expressions of the form "purple N" are conventionally used to express ideas of the form "purple N," and so on, we are not making "an arbitrary extrapolation from structural correlations in usage, one that arises from our simplifying proclivities as theorists." For the "correlations" on which we are relying must be *conventions*. Conventional regularities are more than mere accidental correlations, as we discussed in §9.3. I suspect that it simply did not occur to Loar that the use of expressions of a given structure to express ideas of a given structure could be conventional. The notion fits quite naturally with the general conception of mind and meaning that he developed, though. And given that people have the cognitive capacity to recognize not only particular expressions and ideas, but also their structures, the notion that there are conventions relating the latter is no more far-fetched than the platitude that there are conventions relating the former.

§10.4 IMPLICATURE CONVENTIONS

A problem for any use theory is to distinguish meaning from what H. P. Grice (1975) called "implicature."[10] A speaker *implicates* that p iff he means

10 See Grice 1989 for the most complete exposition of Grice's work on implicature, and Neale 1992 for an excellent survey and assessment. I critically analyze Grice's theory of implicature in Davis 1998, which can be consulted for further references. Neal (1992: 555–6) briefly addresses the problem discussed in this section at the conclusion of his article, in somewhat different terms. Neale wistfully expressed disappointment that Grice made no progress on this "urgent" problem. See also Frege 1892b: 73–4; 1918: 9; Grice 1961: §3; Kempson 1975;

or implies that p without saying that p, and does so by saying something else. Suppose that Ann asks Bob, "Where can I get some gasoline?" Bob might reply, "There is a station around the corner," implying that Ann can get gasoline at the station around the corner. Then he expressed that belief, and meant this in the inclusive but not the exclusive sense of speaker meaning. Bob did not *say* that Ann can get gasoline at the station around the corner, because that is not what the sentence he used meant. So Bob "implicated" this. Figurative usage is another case in which implicatures arise. Speakers who use "That's great!" ironically implicate, but do not say, that something is the very opposite of great.

Grice went on to distinguish *conversational* from *conventional* implicatures. Bob's implicature is conversational, because it is generated in some way from the conversational context, not just from the meanings of the words he used. Bob could have used "There is a station around the corner" without having implied that Ann could get gasoline there. This would not be a cooperative thing for Bob to do, since he would thereby mislead Ann into thinking that she could. But Bob would be using proper English; he would not have made any linguistic mistake. Moreover, he could perfectly well have said "There is a station around the corner, but it is closed," in which case he would not even have implicated that Ann could get gasoline there. Grice would say that the but-clause *canceled* the implicature. By contrast, someone who says "Bill is sick, so he should rest" normally implicates that Bill's being sick implies that he should rest. This implicature is said to be conventional, because it arises in some way from the conventional meaning of "so," and cannot be canceled. A speaker who says "Bill is sick, so he should rest, but his being sick does not imply that he should rest" has not exactly contradicted himself. But his but-clause undermines his use of "so."

As in the case of "means," "expresses," and "implies," "implicates" can be applied to *sentences* as well as to *speakers*. A sentence implicates, roughly, what speakers using the sentence with its regular meaning would commonly use it to implicate. Despite what Bob implicated in his response to Ann, the sentence Bob used does not itself implicate "Ann can get gasoline at the station." In many contexts, the use of "There's a station around the corner" would not suggest that the hearer can get gasoline

Levinson 1983: §3.2.3; Rieber 1997; and especially Bach 1999. Bach tries to show that what Grice and others call conventional implicatures are really part of what is said. Kempson argues that they should be classified as presuppositions, entailments, or conversational implicatures, while Kartunnen and Peters 1979 argue that what we usually call presuppositions are better classified as conventional implicatures.

there, and its use would not be misleading if the speaker thought the station had no gasoline, or was closed. Indeed, the speaker may have been talking about a train station. In contrast, (5)(a) itself implicates (5)(b).

(5) (a) Bill is sick, so he should rest.
(b) Bill's being sick implies that he should rest.

A speaker could not properly use (5)(a) without implicating (5)(b). And (6)(a) implicates (6)(b):

(6) (a) Some died.
(b) Not all died.

For speakers would not normally use "Some died" without implicating "Not all died." In the case of (6), the implicature is conversational, and can be canceled. "Some died, but not all did" is not contradictory or linguistically odd in any way.

I have argued elsewhere (Davis 1998) that even conversational implicatures depend on conventions when we are talking about sentence rather than speaker implicatures. The sentence form "Some S are P" is commonly used to implicate that not all S are P, and this practice meets all of the conditions for something to be conventional. The practice is not universal, but it is certainly common: speakers regularly use sentences of the form "Some S are P" to implicate that not all S are P, and are regularly so interpreted. The regularity is socially useful, contributing as it does to communication. Speakers who implicate that not all S are P communicate the thought that not all S are P when their implication is recognized. The regularity is self-perpetuating: the fact that speakers before us have used "Some S are P" to implicate "Not all S are P" gives us a good reason to do so ourselves; the habit of using the sentence to express the thought, and the association between the two, are both reinforced each time "Some S are P" is used to implicate "Not all S are P"; and the practice is transmitted from one generation to the next as language is learned. Lastly, the regularity is arbitrary. We could have used "Some S are P" to implicate "It is not the case that 50 percent of all S are P" or "Perhaps all S are P" or "It is an open question whether all S are P." Or we could have used "Some S are P" with no implicature at all about stronger propositions, the way we use "At least 99 percent of all S are P." The arbitrariness of conversational implicature conventions conflicts with the Gricean assumption that conversational implicatures are "calculable" from and "determined" by contextual factors together with general psychosocial principles, namely, the cooperative principle and

associated maxims. I have argued directly against these claims in the work cited.

The conventionality of sentence implicatures creates the following problem for the theory that word meaning is the conventional expression of an idea or other mental state. Given the implicature convention just noted, "Some S are P" is conventionally used to express the belief, and therefore the thought, that not all S are P. So why doesn't "Some S are P" *mean* "Not all S are P"? The fact that "Some S are P" is not always used with its customary implicature is not the answer. For the usage conventions underlying word meaning are almost invariably nonuniversal (§9.5). The answer, I believe, lies in the *indirection* that is characteristic of implicature. Speakers who use "Some died" to express the thought that not all died do so *by* expressing the thought that some died. Hence even though they use "Some died" to mean in the inclusive sense that not all died, they do not mean "Not all died" by the sentence "Some died." We have an instance of *cognitive* speaker meaning here, but not *cogitative*; *expression* but not *direct expression*. In moving from the simple to the refined Gricean analysis of word meaning, we moved from using cognitive speaker meaning and the expression of belief to using cogitative speaker meaning and the direct expression of thought. We did so in order to rule out figurative usage and in order to rule in fictional uses, indexicals, and subsentential meaning (§8.5). In the process, we correctly excluded conventional implicatures. The special meaning of terms like "so" and "but" that generate noncancelable implicatures will be discussed in my *Nondescriptive Meaning and Reference*.

§10.5 NONSPECIFIC CONVENTIONS

The referential-attributive distinction will help to illustrate a subtle feature of the recursive neo-Gricean analysis that might appear problematic. As Loar (1976a: 376) observed, the referential use of definite descriptions is just as conventional as the attributive (see §6.1). That is, people regularly use expressions of the form "the F" to refer to some individual N, where "N" is a proper name or pronoun, and this regularity has the other features of a convention. Hence it is conventional to use "the F" to express some idea other than the idea "the F." Because of this convention, sentences of the form "The F is G" are conventionally used to implicate some proposition of the form "N is G" (§10.4). Since "the F" does not mean "N" in English, even though it may be used to refer to N, the neo-Gricean analysis

may appear to be too weak. However, there are two reasons why the neo-Gricean analysis rules that the referential use of a definite description is not part of its meaning. First, we use "the F" to express the idea of N *indirectly*. "Mt. Everest is in Nepal" is at best a conventional conversational *implicature* of "The tallest mountain is in Nepal" (§10.4). By contrast, when what may appear to be a description is used as a name, there is no such indirection. "The United States" is conventionally used as a name of America, the North American country whose capital is Washington. But it does not express the idea of America by expressing the idea of the contextually determinate states that are united. It may have done so in 1789, but no longer. So "the United States" means "America," not "the contextually determinate united states." Second, while it is conventional to use expressions of the form "the F" to refer to *some* individual N, there is generally no *particular* individual N to which "the F" is conventionally used to refer. The convention is *nonspecific*. While I might use "The woman I married" to refer to Kathy Olesko, it is not conventional for people to use "The woman I married" to refer to Kathy Olesko. "The tallest mountain" is exceptional in this regard.

Many noun compounds are governed by an equally nonspecific convention. Some have been lexicalized, with a determinate meaning accounted for by the base clause of the neo-Gricean analysis. Examples are *apple pie*, *desktop computer*, and *car stereo*. Some are instances of conventionally used forms, with meanings accounted for by the recursion clause: *coffee lover, cop killer,* and *Russian teacher*. But there is, in addition, a nonspecific convention to use a compound of two nouns "N_1" and "N_2" creatively to express a contextually definite idea containing in some way the ideas expressed by "N_1" and "N_2." Thus a compound like *Ferrari woman* might be used to mean a woman who drives a Ferrari, a woman who works for the Ferrari company, a woman who is a member of the Ferrari family, a woman who is standing under a Ferrari sign, a woman who had a scandalous relationship with a member of the Ferrari family, and so on endlessly (H. H. Clark 1983: 311, 326). While it is conventional to use a noun compound "$N_1 + N_2$" to express the idea of an N_2 with some contextually definite relation to N_1, there is no specific relation R such that "$N_1 + N_2$" is conventionally used to express the idea of an N_2 related by R to N_1. So even though there is no indirection, the neo-Gricean analysis rules correctly that a noun compound like *Ferrari woman* does not *mean* "woman who drives a Ferrari" (or any of the other possibilities just mentioned) in English.

Predicates of the form "finished N" illustrate an even more complex interplay of the base and recursion clauses. We observed earlier that the two clauses sometimes assign the very same meaning to a compound (e.g., "cat lover"), while to others they assign totally distinct meanings (e.g., "kick the bucket"). In some cases it is not obvious whether we have a single general compositional meaning or several specific lexicalized meanings. Consider "finished the book." It is initially tempting to say that this phrase is ambiguous, with at least two meanings: "finished reading the book" and "finished writing the book." But a little reflection leads to the realization that there is really no limit to what we could use "finished the book" to mean. We could use it to refer to the completion of any of the following processes: printing the book, binding it, burning it, painting it, weighing it, wrapping it, chemically analyzing it, assessing its likely market value, and so on ad infinitum. Furthermore, "finished dinner" most commonly means "finished eating dinner" or "finished preparing dinner," but it could also mean finished serving, photographing, or analyzing dinner.

The next natural thought is that "finished N" has a single indexical meaning,[11] something like "completed the contextually indicated process involving N." Speakers will of course also use the phrase to mean something specific, expressing the purely descriptive idea by expressing the indexical. If the indicated process is washing, then the phrase will be interpreted as conveying "finished washing N." This indexical account does not seem to capture the whole story, however. There are respects in which the use of "finished the book" to mean "finished reading (or writing) the book," and the use of "finished dinner" to mean "finished eating (or preparing) dinner," are special. First, we can use these phrases with those meanings without ever indicating by any other means what process we had in mind. And our audience will generally try to figure out which of the two common meanings was intended, without giving the impression that the range of options is huge. If I say out of the blue, "I finished the book," my audience is more likely to respond "Which book?" than to ask "Reading it or writing it?" And both responses are much more likely than "What were you doing to the book?" Second, these

11 Cf. the discussion of incomplete descriptions in *Nondescriptive Meaning and Reference*. Compare and contrast Recanati 1995: 209–10. Recanati thought that the meaning of "finished N" constitutes an objection to compositionality. As we will see in §10.6, his objection was based on inexplicably assuming something about the linguistic meaning of "finished N" that he denies about the meaning of "finished."

meanings are often intended and understood even when other actions are contextually indicated. Suppose I said, "Johnny played with his dinner for a long time. He finished dinner at 7:30." I could have meant that Johnny finished playing with his dinner at 7:30, but I doubt that anyone would interpret me as meaning that. I am more likely to have meant, and to be understood as having meant, that Johnny finished eating dinner at 7:30 (which would imply that he finished playing with his dinner somewhat earlier). By contrast, the unambiguously indexical "He stopped doing it at 7:30" would inevitably be interpreted as saying that Johnny stopped playing then. It is for reasons like these, presumably, that "completed reading" and "completed consuming" are listed as separate meanings for "finished" in some large dictionaries.[12]

So, all things considered, I believe that "finished the book" and "finished dinner" do have a few specific meanings in addition to their general indexical meaning. Substitution instances of "finished N" get their general indexical meaning from clause (ii). "Finished dinner" and "finished the book" get their specific meanings from clause (i). Whereas "finished the book" is conventionally used to directly mean "finished reading the book" or "finished writing the book," its use to mean "finished burning the book" or "finished photographing the book" are too unusual and context-dependent to qualify as conventional. In particular, while it seems accurate to say that we use "finished the book" to mean "finished reading the book" on a given occasion because people have done so before, it does not seem accurate to say that we would ever use "finished the book" to mean "finished photographing the book" because people have done so before (see §9.3). The precedent that we would be following is the generic indexical convention.

§10.6 OBJECTIONS TO COMPOSITIONALITY

The recursion clause of the neo-Gricean analysis is based on the compositionality of linguistic meaning. As axiomatic as compositionality has seemed to most students of language, it has recently come under criticism.[13] The first line of objection is based on the way that context

12 For example, the *New Shorter Oxford English Dictionary*.

13 See, e.g., Kittay 1987: Chapter 3; Ruhl 1989: Chapter 1; Lahav 1989; Grandy 1990: 568; 1998: 22; Recanati 1995: 209–11. H. H. Clark (1983, 1993) provides relevant background information about the role of context in the interpretation process, although he does not argue for or against compositionality.

influences our understanding of sentences. If the objections undermine compositionality, then the recursive neo-Gricean analysis ignores some determinants of complex expression meaning. Consider:

(7) He banked the money.
(8) He banked the plane.
(9) He banked too little.

It might be argued that "banked" must mean "deposited in a bank account" in (7) and "rolled into a banked orientation" in (8). The first two sentences are not ambiguous in the way that (9) is. It thus appears in this case that the meaning of a component is determined by the compound that it is contained in, thereby falsifying compositionality, which claims that the meaning of a compound is determined by the meanings of its components. If this line of reasoning were correct, then the recursion clause in the neo-Gricean analysis would have to be restricted in some way to prevent assignment of the meaning "He rolled the money away from horizontal" to (7) and the meaning "He put the plane in the moneylending institution" to (8).

The reasoning is incorrect for two reasons. First, the most we could conclude is that the meaning of one component of a compound sometimes determines which meaning of another component is operative. Even if the meaning of "the money" determines that "banked" means "deposited in a bank account" in (7), the meaning of the compound "banked the money" would still be determined by the meaning of "the money" and the selected meaning of "banked" therein. (*Analogy*: If an electorate is polarized in such a way that one of two subgroups always votes opposite the other, the overall vote will still be the resultant of the votes of the two groups even though the vote of one group is determined by the vote of the other group.) Second, as we noted in connection with (1), sentences (7) and (8) *are* ambiguous. "He banked the money" can be interpreted as meaning "He rolled the money into a nonhorizontal orientation." We can fancifully imagine a man making a plane out of stacks of money, and then flying it away to avoid the police, during which flight he inevitably turned several times. Then he banked the money. To imagine "He banked the plane" true when "banked" means "deposited," all we need to imagine is a bank that takes planes on deposit, lending the planes for a fee that enables the bank to pay depositors interest and still make a profit. We never hear these meanings of (7) and (8) in ordinary contexts because we know that money cannot fly and that there are in fact no plane banks, and it normally does not occur to us that speakers

might have such unrealistic scenarios in mind. Sentence (9), "He banked too little," is ambiguous in a way that (7) and (8) aren't, because nothing in the linguistic context in which "banked" occurs in (9) provides a clue to its intended meaning. The nonlinguistic context must be relied on to determine what the speaker means. Is the man flying a plane, or handling money?

What (7) and (8) illustrate is a problem not for a semantic theory of meaning but for a cognitive theory of understanding. Why and how do speakers of English in ordinary contexts automatically understand (7) as meaning "He put the money in a bank" and (8) as "He put the plane into a banked orientation"? How do they determine which meaning of (9) is intended on any given occasion? How do hearers determine when the speaker means something by an expression that the expression does not mean in the language being used? It seems clear that hearers automatically apply their general knowledge of planes and money, along with their knowledge of the particular speaker and context, and normally do not entertain completely unlikely scenarios. While the meaning of an expression is determined by the concept it expresses, the hearer's interpretation of an expression is determined by her conception of its referent and the context (§19.4). The hearer's knowledge of the language is also essential, of course. The hearer might conclude that (9) means "He did not build up the banks enough" if it occurs to her that "bank" can mean "build up banks (as of snow or dirt)" in English, but not if she ignores or is ignorant of this particular meaning of that highly ambiguous word. The problem of understanding understanding is closely related to the problem in psycholinguistics and artificial intelligence of developing a "parser" – an artificial system or program that analyzes speech and determines what the speaker meant. Advances in one area suggest leads for the other. A theory of understanding is at least as important for cognitive science as a theory of meaning. But the phenomena to be explained by the two theories, though obviously related, are different.

Recanati claims that "the conventional sense of the constituents of a complex phrase and the way they are grammatically combined is not sufficient to determine the semantic value of the complex phrase involving those constituents" (1995: 209). If this is a claim about semantic determination and not about cognitive interpretation, then it is a denial of compositionality and contradicts the recursive neo-Gricean analysis. Recanati tries to show that context can do more than select one of the conventional meanings of an expression to be operative in a compound.

He cites cases in which a general term receives a specific interpretation in context. Consider:

(10) She wears rabbit.
(11) She eats rabbit.
(12) She wants rabbit.

Recanati makes two claims. The first is innocuous: "rabbit" would be interpreted as meaning "rabbit fur" in (10) and "rabbit meat" in (11). The second is questionable: these are not conventional meanings of "rabbit." The second claim is false, I believe, because speakers commonly use "rabbit" to mean "rabbit fur" in many contexts such as (10) and (12) and "rabbit meat" in many other contexts such as (11) and (12), and these common practices have all the marks of a convention (see Chapter 9). These meanings are listed separately in some dictionaries.[14] And (12) seems obviously ambiguous. These uses of "rabbit" are very different from a case in which someone uses it to mean "rabbit lymph" in "She drank rabbit." *That* would be an unconventional use of "rabbit," as easy as it might be to figure out what the speaker meant in an appropriate context.

Our current focus is on the recursion clause of the neo-Gricean analysis. Recanati's second claim is a denial of the *base* clause. To get from Recanati's claims to a denial of compositionality or the recursion clause, we need a third premise: in English, the phrase "wears rabbit" has the specific meaning "wears rabbit fur," and "eats rabbit" has the specific meaning "eats rabbit meat." This premise is correct, I believe, but it undercuts the second premise. No reason offered or imaginable supports the claim that "wears rabbit" means "wears rabbit fur" in English while "rabbit" does not mean "rabbit fur" in English. If "rabbit" has only a single generic sense even though speakers intend more specific interpretations in context, then it should have only a single generic sense in compound expressions of the form "Vs rabbit," even though speakers may give them more specific interpretations in particular contexts. If "rabbit" applies to both rabbit fur and rabbit meat in some common neutral sense, then "wears rabbit" should apply to both people who wear rabbit fur and people who wear rabbit meat in the same neutral sense.

Recanati cites Bach (1994a), who uses Grice's razor to argue against the postulation of multiple senses in many different cases. But if Grice's

14 The *New Shorter Oxford English Dictionary* lists the meat sense separately, and the *Webster's Third New International* lists the fur sense separately.

razor leads to the conclusion that "rabbit" has only one generic meaning, then it should lead just as well to the conclusion that it has only that meaning in compounds.[15] Bach in turn cites Ruhl (1989), who adopts the "monosemic bias" as a methodological principle, attributing most apparent ambiguity to interactions of a general, abstract meaning with a specific linguistic or nonlinguistic context. Ruhl overlooks the fact that a term like "rabbit" can be interpreted in such a way that it is false of rabbit meat, or false of rabbit fur, as we choose. We can, at will, interpret "wears rabbit" so that it means "wears rabbit meat" (and then is false of someone wearing rabbit fur) or "wears rabbit fur" (and then is false of someone wearing rabbit meat). If "rabbit" were truly general in meaning rather than ambiguous, then it would not have a sense in which it is false of rabbit fur or false of rabbit meat. More importantly, if the monosemic bias leads us to conclude that "rabbit" is general rather than ambiguous, then it should also lead us to conclude that "wears rabbit" is general rather than ambiguous. So if they are consistently applied, neither Grice's razor nor the monosemic bias should lead to a rejection of compositionality.[16]

Lahav (1989) has argued against compositionality on the basis of the noun dependence of adjectives. Consider:

(13) This is a red wall [its surface is red].
(14) This is a red apple [its skin is red].
(15) This is a red grapefruit [its meat is red].
(16) This is a red pen [its ink is red].

"Red" normally does mean different things when it is combined with different nouns, as these examples and many more illustrate. But that does not contradict compositionality. For in every case, the meaning of the compound "red N" is determined by the meanings of "red" and "N" *in that compound*. In "red wall," "wall" means "wall," and "red" means "having a red surface"; hence "red wall" means "wall with a red surface." In "red grapefruit," "grapefruit" means "grapefruit" and "red" means "having red meat"; hence "red grapefruit" means "grapefruit with red meat." The fact that the meaning of one term in a compound varies with the meaning of another term does not imply that the meaning of the

15 See Davis 1998 for a critical discussion of Grice's Razor.

16 Travis (1997) would argue that the word has no fixed meaning at all. But his argument depends on the false assumption that a word means something on an occasion if a speaker uses it with that meaning (contrast §7.9), and on the fact that speakers can use vague terms with an indefinite number of more precise meanings. Travis also seems to assume that vague truth conditions are not truth conditions.

compound is not composed of the meanings of its components. Lahav inferred noncompositionality from noun dependence because he took compositionality to require that a component has the same meaning in all of the compounds it appears in.[17] But compositionality need not be linked with the assumption that the meaning of a word must be uniform in this way.

Lahav's argument against compositionality is the reverse of Cruse's (1986: 37) and Cowie's (1994: 3169) argument against the traditional definition of idioms. According to most dictionaries, an idiom is "an expression whose meaning cannot be predicted from the meanings of its parts." The objection is that by "the meanings of its parts" we must mean the meanings of its parts *outside of idioms*, rendering the traditional definition circular. But this objection is groundless. Consider the idiom "raining cats and dogs," which means "raining hard." We cannot infer from this that "cats and dogs" means "hard" in this idiom. "Cats and dogs" does not mean "hard" in English, and speakers do not mean "hard" by "cats and dogs" when they use "raining cats and dogs" to mean "raining hard." Similarly, "kick" and "bucket" do not have meanings that could combine to form the idiomatic meaning of "kick the bucket." And speakers who use "kick the bucket" to mean "die" do not mean anything by "kick" and "bucket" on such occasions, except perhaps "kick" and "bucket."

The statement that e "normally means" μ in certain compounds is a generalization about how the expression is actually used. It concerns *applied* word meaning rather than *linguistic* word meaning (§7.9). Thus "red" can mean "having a red surface" in some compound even though it normally means "having red meat" there. Neither compositionality nor the neo-Gricean analysis requires that applied word meaning be uniform across contexts. The recursion clause of the neo-Gricean analysis as it is formulated, however, does assume the uniformity of linguistic word meaning. It assumes that *a word can have any of its meanings in all of the compounds it appears in*.[18] Hence every ambiguity in a word will produce a corresponding ambiguity in any compound containing it. Let us call this the *semantic universality* assumption. I believe authors like Katz and Fodor (1963: 493) who deny the semantic universality assumption do so because they take the goal of a semantic theory to be an account of the

17 Lahav 1989: 262. See also Fodor and Pylyshyn (1988: 124–5), who are overreacting to Smolensky's (1988: 66–7) clear denial of compositionality.

18 Holding constant grammatical category. Words like "fall" occur in different grammatical categories, with different meanings. "Fall" will obviously never have its noun meaning in a compound in which it appears as a verb. See §10.7.

speaker-hearer's ability to understand sentences, thus conflating a theory of meaning with a theory of understanding.

The noun dependence of adjectives would be a problem for the recursion clause of the neo-Gricean analysis if adjectives had some conventional senses that they did not have in all compounds in which they appear. I accept the semantic universality assumption, and so find the recursion clause acceptable as it stands. I believe that if a word has a sense, that sense will be part of at least one meaning of any compound it appears in. That interpretation of the compound may be "unnatural," however, and very difficult to "hear." These facts, I believe, are due to facts about the cognition of understanding, not about the language. To see this in the earlier examples, note first that there is an interpretation of (13)–(16) in which "red" has a uniform meaning – the meaning it has in (13). An apple, grapefruit, or pen would be red, in a sense, if it were painted red. There is even a sense in which a green apple would be red if, surprisingly, its meat were red, and in which a pink grapefruit would be red if its peel were red. The apparently contradictory sentence "That red pen is not red" might be true, in one sense, because its ink is red while its case is blue.[19] Some of these interpretations are more natural and consequently easier to hear, that is, to process cognitively so that they register in linguistic intuition. Since apples are known to vary widely in skin color while having nearly uniform meat color, and are never painted, we naturally interpret a sentence like (14) as being about skin color. Since grapefruit are known to vary markedly in meat color but only slightly in skin color, and are never painted, we naturally interpret a sentence like (15) as being about meat color. Our cognitive processing mechanisms are tuned by experience to interpret language in an efficient manner.

Some meanings of compounds are almost impossible to hear because on those interpretations the compounds are nonsensical. Thus it is very difficult to hear (15) as saying that the ink of the grapefruit is red, or to hear (16) as saying that the meat of the pen is red. For we know that it would be absurd to say that a grapefruit has ink or that a pen has meat. It would take a Walt Disney or Lewis Carroll fantasy to imagine someone writing with a grapefruit, or breaking open a pen to get at the meat

19 "Red N" thus differs markedly from "good N," which has a single general meaning "good for an N." If we give someone a pen as a present, it may be a poor present even if it is a good pen. But then there is no sense in which it is a good present but a poor pen. Compare and contrast Lahav 1989: 267–70. "Red N" also differs from "finished N," to be discussed, which has a general indexical sense as well as specialized descriptive senses for particular choices of "N."

inside. In such fantasies, (15) or (16) might be true in the sense indicated. I would be willing to bet that no one who has read this far has thought of any of these sentences as saying that the wall, apple, grapefruit, or pen is communist, even though "communist" is one of the senses of "red." We can interpret the sentences as saying that the objects in question are communist, but the claims would then be so absurd that the thought expressed would not have crossed any speaker's mind in normal contexts. We can imagine communist grapefruits and pens, however, by visualizing a Disney cartoon in which Mickey Mouse reviews columns of grapefruits and pens marching by in Soviet uniforms during a May Day parade. With that scene in mind, we should be able to hear "red" as meaning "communist" even when applied to grapefruits and pens. We do not need flights of fancy to hear the anomalous interpretations of negations of (15) and (16). It suffices to recall or formulate reasons for thinking that (15) or (16) cannot possibly be true in the anomalous sense, and then to construct an argument. We should be able to interpret "This is not a red grapefruit because fruit cannot be communist" in such a way that "red" means "communist" and the argument is sound (see §10.8). (We should also be able to hear the argument as a non sequitur, with "red" having its normal grapefruit sense.)

Kittay (1987) seems to argue that the pervasiveness of polysemy provides a reason to reject compositionality. Doing some quick math, she concludes that the number of meanings that even short sentences have would be "absurdly numerous" if compositionality were true, given that most words have several senses. Each word other than the article in the following sentence, for example, has at least five different senses.

(17) The slow red rabbit finished the red bank.
5 5 5 5 5 5

Since the sentence contains six words with at least five senses, Kittay takes compositionality to imply that it has at least $5^6 = 15{,}625$ senses. Katz (1972: 92) too thought it would be absurd for the number of senses to be even a fraction of that, so he imposed restrictions on the operation of projection rules. Since compositionality still holds on Katz's theory, we see that the objection is not to compositionality but to an unrestricted recursion clause. Every meaning of a compound may be determined by the meanings that its components have in that compound, even if the compound lacks some meanings that might be predicted from the meanings of the components. Interactions among the meanings of the components may block certain meanings of the compound.

Now 15,625 is truly an amazing number of senses, especially since we do not normally "hear" more than one or two senses at any given time. But while the size of this number may be surprising, it is hard to find any reason for thinking that the number is absurd. It would be absurd to claim that any speaker had all of these senses in mind on any given occasion, because our mental capacities are severely limited in this respect. It would also be absurd to claim that any expression had so many meanings on a particular occasion. For that would imply that the speaker meant all of those things, which is also impossible. But the fact that a sentence has a meaning in English does not have such implications concerning individual speakers. Indeed, the size of the number just seems to be another demonstration of the remarkable productivity of natural languages. Given that the categories of adjectives, nouns, and verbs each contain over a thousand words in English, it is easy to see that English must contain over 1000^6 = one quintillion sentences with the same structure as (17). This means that each speaker of English is capable of understanding an astronomical number of sentences. That is remarkable, even mind-boggling, but not absurd.

It could never be true, of course, that anyone had actually understood all of the sentences with the same structure as (17). But each one of the sentences is such that a competent speaker could easily understand it. We can verify this claim by randomly generating sentences with the same structure and observing that each one is easily understood: "The big blue fox ate the soft rabbit," "The small hot rock broke the expensive window," and so on. In a similar way, we can verify the claim that (17) has all of those senses by observing that for each one we can construct a linguistic context in which that is the sense we hear. For example, we are very unlikely to hear "finish" in (17) as having the sense that it has in "Bob finished the race." But we are almost forced to hear it with that sense in (18) and (19):

(18) *Because commercial banks are not races and thus cannot be finished, it cannot be true that* the slow red rabbit finished the red bank.

(19) *If the bunny rabbit with red fur that cannot run fast completes every race it runs and the savings bank painted red is a race, then* the slow red rabbit finished the red bank.

For each one of the senses of (17), we can fill in the blank in (18) and (19) in such a way that the occurrence of (17) in (18) or (19) is naturally understood in that sense. A speaker who used the sentence with that meaning would be speaking perfectly correct English. A description of

English that does not in some way assign that meaning to (17) must therefore be incomplete or false.

Compositionality does amplify the theoretical problem to be solved by a cognitive theory of understanding. In addition to explaining how people have the ability to understand an astronomical number of different sentences, each with a large number of senses, cognitive theory must explain how people arrive at a limited number of plausible interpretations of a sentence like (17) with little or no conscious effort in no more than a second. The large number of meanings that (17) has rules out some possible explanations. The normal process of understanding (17) cannot involve explicitly representing each meaning of the sentence, and then systematically ruling them out one by one until only the most plausible remain. We just cannot think fast enough to do that in a second. But this overly intellectualist hypothesis is implausible for independent reasons, and is hardly the only possibility. An explanation along connectionist lines seems much more promising (see Chapter 18). Whereas the idea "wall" primes the idea "having a red surface" and inhibits the idea "having red meat," the idea "grapefruit" primes the idea "having red meat" and inhibits "having a red surface." These facts about the association of ideas at least partly explain why we automatically interpret "This is a red wall" and "This is a red grapefruit" in the way that we do.

If I am wrong in thinking that a word has all of its conventional meanings in any compound in which it appears, that would falsify the recursion clause of the neo-Gricean analysis as I have formulated it. But the clause could be reformulated without abandoning the general idea that word meaning is constructive conventional speaker meaning. In the next section, we will restrict the recursion clause so that it only generates meanings for *grammatical* compounds. I assume that ungrammatical expressions like "John are nicely" have no meaning in English. We could strengthen this restriction by requiring that e be a *well-formed* expression, where there are semantic as well as grammatical constraints on well-formedness. Alternatively, we could follow Katz (1972: 92–3) in imposing *selection restrictions* on the operations of the projection rules underlying the recursion clause, in order to ensure that the senses they generate for linguistic compounds on the basis of the senses of their components and mode of combination do not combine incompatible senses. The meanings assigned to "A" and "B" by the base clause of the neo-Gricean analysis could conceivably be incompatible not just in the familiar sense that "A" and "B" could not both be true of anything, but in the distinct sense that "A" and "B" could not both have those meanings when appearing together in certain

compounds. I do not know of any grammatically combinable expressions that are semantically incombinable, so I do not believe that there are any purely semantic constraints on well-formedness. Katzian selection restrictions in fact rule out senses that compounds actually possess. We will return to this issue in §10.8 when we discuss anomaly.

§10.7 THE GRAMMATICALITY RESTRICTION

We have defended the recursive neo-Gricean analysis against several objections, but genuine problems remain. Our current formulation requires three revisions. First, the recursion clause of (3) and (4) assigns meanings to some ungrammatical compounds. It is conventional to use expressions of the form "A-er" to mean "more A," or more properly, "greater in A-ness," as in "higher," "faster," and "farther." It is conventional to use "A-est" to mean "most A," as in "highest," "fastest," and "farthest." But then the recursion clause (ii) as formulated in (3) and (4) certifies the childhood declension of "good": *good, gooder, goodest*. That is, (3) and (4) imply that "gooder" means "greater in goodness" in English because "gooder" is an expression of the form "A-er," and "A-er" is conventionally used to mean "greater in A-ness." But while we know how to interpret these expressions when someone mistakenly uses them, "gooder" and "goodest" are not grammatical English words, and so do not mean anything in English.

We can avoid this problem by requiring that $E[e_1, e_2, \ldots, e_n]$ be *grammatical* – that it conform to the syntactical rules of the language under discussion. All of the morphemes of the language are grammatical, but not all expressions formed from the morphemes are grammatical. Conventions to use a form will always be conventions to use grammatical instances of the form. A substitution instance of the form will get a meaning from the conventional use of the form only if the substitution results in a grammatical instance of that form. The proposed addition to (3) is italicized in (20).

(20) Expression e expresses i iff (i) people conventionally use e to directly express i, or (ii) people conventionally use $E[x_1, x_2, \ldots, x_n]$ to directly express $I[i(x_1), i(x_2), \ldots, i(x_n)]$, *where* $e = E[e_1, e_2, \ldots, e_n]$ *is grammatical*, $i = I[i(e_1), i(e_2), \ldots, i(e_n)]$, and $i(e_i)$ is the idea expressed by e_i for all i from 1 to n.

Thus revised, the neo-Gricean analysis yields the correct result that "gooder" *would* mean "better" if it *were* grammatical, without producing the result that "gooder" *does* mean anything in English. No grammaticality restriction is needed on the base clause. Conventional usage of a letter or

sound sequence, no matter how irregular, makes it an expression of the language and gives it a meaning. The fact that "alumni" is an English plural noun and "ad hoc" an English adjective are good examples.

Lehrer (1974: 96, 197) and Allerton (1987: 66) note that word formation, as opposed to phrase, clause, and sentence formation, is generally irregular, with defective patterning. Pawley (1985: 101, 110, 114) labels it a "non-productive" process.[20] There are many cases, furthermore, in which it is unclear whether a complex expression is an English word or not. Thus Landau (1984: 78) complains that dictionaries often inflate their word counts (a key marketing feature) by including words that could exist but for which no record of use actually exists. He cites "oppressingly," "sluggardliness," and "idioticalness." To the extent that it is unclear whether these are words in English, it is also unclear whether they have a meaning in English. It is no strike against the neo-Gricean analysis that it does not solve this problem for syntactic theory. We can rest content in the knowledge that the neo-Gricean theory correctly predicts what the words would mean if they were grammatical.

The *gooder* example is unusual in several respects. Let E_{comp}[ADJ] be the function representing the form "A-er." E_{comp}[ADJ] takes an adjective "A" and produces the comparative form "A-er" in the "regular" way, by adding "er." This form is unusual in one respect because it has substitution instances that are grammatical (e.g., *tall, taller*) as well as substitution instances that are ungrammatical (e.g., *good, gooder*).[21] More seriously ungrammatical forms have no grammatical instances at all. Consider the form of "The of boy if well": ART + PREP + NOUN + CONJ + ADV. Whenever a form has no grammatical substitution instances, there will be no convention to use it to express anything. There is no idea structure conventionally associated with the linguistic form. So even without the grammaticality restriction, the recursion clauses of (3) and (4) would not assign a meaning to "The of boy if well."

The *gooder* example is unusual in two further respects: it involves compound words rather than compounds of words; and we know what idea the ungrammatical compounds would express if they were grammatical. Consider "John are bald," which is ungrammatical because the subject and predicate do not agree in number. This sentence consists of a subject

20 See O'Grady et al. 1993 for an introduction to morphology.

21 Analogy: in propositional logic or quantification theory, a logical form that is contingent (e.g., p&q) will nevertheless have some substitution instances that are logically true (e.g., [Av–A]&[Bv–B]) and some that are logically false (e.g., A&–A) as well as some that are contingent (e.g., A&B).

followed by a predicate, and so has the form E_{sp}[NP, VP] = NP + VP. It is conventional to use sentences of the form NP + VP to express subject-predicate thoughts. People use "The sky is blue" to express the thought "The sky is blue"; they use "Red is a color" to express the thought "Red is a color"; and so on. In general, expressions of the form E_{sp}[NP, VP] are conventionally used to express ideas of the form I_{sp}[i(NP), i(VP)], which are thoughts combining i(NP) as subject concept with i(VP) as predicate concept. While there is an idea structure conventionally associated with subject-predicate sentences, it is not clear whether the ideas "John" and "are bald" can be combined in the indicated way. That is, it is unclear whether I_{sp}[i(NP), i(VP)] is defined for the ideas "John" and "are bald." We know that there is no such thing as the idea "John are bald" Since "John are bald" is ungrammatical and does not express any idea in English, "the idea 'John are bald'" is ungrammatical and does not refer to anything (Definition 7.5). This does not entail, though, that there is no thought whose subject concept is the idea expressed by the singular subject "John" and whose predicate is the idea expressed by the plural predicate "are bald." I_{sp}[i(*John*), i(*are bald*)] might exist, even though we cannot refer to it in the usual way. If we had an argument that there is no such thought, then we could combine it with the recursion clauses of (3) and (4) to rule that "John are bald" has no meaning, even without the grammaticality restriction. But in the absence of such an argument, we need to rely on the restriction. We have to rely on the grammaticality restriction in the case of *gooder* for a more powerful reason: we know that the corresponding idea exists. We cannot refer to it as the idea "gooder," or as the idea of being gooder. But since the idea is expressed by "better," we can refer to the idea as the idea "better," or as the idea of being better.

The recursive neo-Gricean analysis implies that the meaningfulness of a compound depends on its grammaticality. The grammaticality of a compound in turn often depends on or varies with the meanings of its components. Consider "John fell hard," which has the form "N + V + Adv." This form is conventionally used to express thoughts of a familiar structure, such as the thought that time passed quickly. The verb "fell" is conventionally used to mean two different things: "fell down" and "cut down." Without a restriction excluding ungrammatical compounds, the recursion clause of the neo-Gricean analysis might assign two meanings to "John fell hard," one for each of the meanings of "fell." But the sentence is not ambiguous in this way. "Fell" cannot mean "cut down" in "John fell hard." If it did, the sentence would be ungrammatical. This example

is similar to the "red pen" and "red grapefruit" examples discussed in §10.6 in that we automatically ignore certain nonsensical combinations of meanings in our everyday processes of interpretation. The examples differ in that the combinations are grammatically permissible in the "red pen" and "red grapefruit" cases but grammatically impermissible in the case of "fell hard." The fact that grammaticality depends on meaning in the way that we have illustrated does not make the revised neo-Gricean analysis circular in any way, due to its recursive character. What we have observed is that the grammaticality (and therefore meaningfulness) of a compound expression may depend on the meanings of its components, not that the grammaticality of a compound depends on what it means.

§10.8 ANOMALOUS SENTENCES

The extension of "grammatical" is not uncontroversial, and may well be somewhat vague and indeterminate. Consider widely discussed examples like the following:[22]

(21) John frightens sincerity.

Sentences like this are said to be *anomalous*. They strike us as odd or ill-formed in ways that self-contradictions ("It is and is not raining") and garden variety lapses of grammar ("Everyone are happy") do not. I believe that (21) is both grammatical and meaningful, although its meaning is nonsensical or absurd (abstract objects aren't the sorts of things that can be frightened). Because its meaning "makes no sense," its use would be odd in most circumstances. Chomsky and others have maintained that (21) is ungrammatical or syntactically deviant on the grounds that the object of the verb "frightens" must be a concrete, animate noun. If this is correct, then the revised neo-Gricean analysis would rule that (21) should be described as having no meaning at all in English. Anomalous sentences present a problem for the neo-Gricean analysis only if there is evidence that they are meaningful despite being ungrammatical, or grammatical but meaningless.

22 See Ryle 1955: 70, 75–6; Chomsky 1961a; 1961b; 1965: 75–9, 157–60; Katz & Fodor 1963: 485, 503, 517–18; Fodor & Katz 1964: 354; Ziff 1964: 392–5; Katz 1964a; 1964b: 527–8; 1972: 5, 43–7, 49, 90–6; McCawley 1968: 124–36; G. Lakoff 1971b; Lehrer 1974: 177–8; J. D. Fodor 1977: §3.5, §5.6; H. H. Clark 1983: 322–3; and Chierchia & McConnell-Ginet 1990: 39–42.

Katz designed his semantic theory with selection restrictions to produce the result that anomalous sentences are meaningless despite being grammatical.[23] I do not believe that this part of Katz's theory was well motivated. In §10.6, I defended the claim that compounds formed from ambiguous terms may have interpretations that it is hard for us to hear because they are nonsensical. "John frightens sincerity" differs from those compounds because it has no conventional interpretation that is not anomalous. We are consequently forced to interpret it in a way that is nonsensical. I believe we nonetheless have to regard (21) as grammatical and meaningful because of the grammaticality and meaningfulness of "frighten" in more complex sentences, such as the following.

(22) Inanimate objects cannot be frightened.
(23) Sincerity is completely outside the realm of things that can be frightened.
(24) I do not believe that anyone can frighten an abstract object like sincerity.
(25) Bill is so foolish he actually believes that John frightens sincerity.
(26) If God can do absolutely everything, then God has to be able to frighten even sincerity, which is absurd, so God cannot be literally omnipotent.

These sentences do not seem odd in any way. The first three are clearly true, and the truth of the fifth is debatable. It is hard to imagine someone believing what (25) says Bill believes. So (25) is unlikely to be true, although it could be. If the sentence embedded in the that-clause were ungrammatical or meaningless, the implied complex sentence containing it would be ungrammatical or meaningless. Thus "Bill is so foolish he actually believes that John frighted sincerity" is as ungrammatical as its subordinate clause.

The fact expressed by (22) (or (23)) is sometimes given as a reason for thinking that (21) is meaningless or ungrammatical. On the contrary, given the syntactic and logical relations between (21) and (22), the truth of the latter entails that (21) is false and thus grammatical and meaningful.[24] Chomsky (1965: 158) observed that the truth and grammaticality of "One cannot elapse a book" does not establish that "John elapsed a book" is either false or grammatical. To be meaningful, I believe that Chomsky's example has to be interpreted metalinguistically, with "elapsed" heard with implicit quotes. What it means is, "One cannot be said to elapse a book," that is, "One cannot say 'Someone elapsed a book." (Compare: "You don't tank people, you thank them," said to correct someone who

23 Katz & Fodor 1963: 485, 503, 517–18; Katz 1964b: 527–8; 1972: 5, 43–47, 49, 90–96.
24 Cf. McCawley 1968: 128.

pronounces "th" as "t.") Chomsky's example cannot be interpreted as meaning that a certain action cannot be performed on a book, because "elapsed" is not a transitive verb and does not denote an action. "Frighten" is a transitive verb and does denote something that can be done to a variety of entities. Hence (22) can be interpreted as saying that that action cannot be done to sincerity. It is the truth of (22) in that sense that shows that (21) is grammatical but false.

I believe that the strongest evidence for the grammaticality and meaningfulness of anomalous sentences, and a clue to their anomalousness, is that we can with a little thought formulate a simple argument that they are false, which we could not do if they were ungrammatical or meaningless. When the argument uses premises and forms of reasoning that every competent language user should know, I refer to it as an *associated refutation*. Argument (27) is an associated refutation for (21).

(27) Abstract objects have no minds.
Objects without minds cannot be frightened.
∴ Abstract objects cannot be frightened.
∴ No one can frighten abstract objects.
Sincerity is an abstract object.
∴ No one can frighten sincerity.
∴ It is impossible that John frightens sincerity.
∴ "John frightens sincerity" is necessarily false.

Another refutation has "inanimate" in place of "abstract." It might be suggested that while *I* am assigning a meaning to (21) in making argument (27), the expression itself does not have a meaning in English. The grammaticality restriction in a definition of linguistic word meaning does not limit which sentences speakers can use to mean something. However, the claim that speaker meaning diverges from word meaning seems indefensible in this case. For the meaning I am assigning to (21) is precisely the meaning you would expect it to have given the meanings of its components in English and the way in which they are combined.

Anomalous sentences with different meanings require different refutations. Consider "The square root of my chair is thirty-five." This is anomalous because we know that chairs are not the sorts of things that have square roots. The proof that chairs cannot have square roots could begin with the premise that chairs are not numbers.

The refutations that we have been examined so far count as proofs because the premises are completely certain and the reasoning self-evident.

When the premises of the associated refutations are less than completely certain, the sentence refuted is not as strongly anomalous. The different ways of completing (28) gradually go from completely nonanomalous to completely anomalous.

(28) John communicated with:
- a. a human.
- b. a chimpanzee.
- c. a dolphin.
- d. a rat.
- e. a snake.
- f. a sponge.
- g. a plant.
- h. a number.

We find the sentences anomalous to the extent that we believe that the kind of object mentioned does not have the mental abilities needed to communicate. In my case, at least, anomaly sets in later in the series with "frightened N" than it does with "communicated with N," reflecting my belief that communication requires more cognitive sophistication than fear. Mere "animacy" is not enough to enable communication. The features that make a compound anomalous do not have the arbitrariness and language-specificity that are characteristic of syntactic phenomena. Anomaly appears to be due instead to what we know about the fundamental nature of the objects that we think and talk about.

In possessing specialized refutations based in some cases on empirical premises, anomalous sentences differ markedly from self-contradictions, which are self-refuting because they violate a basic law of logic. Refutations are neither needed nor possible with genuinely ungrammatical or meaningless sentences. Sentence (21) is very different in this respect from "John frightens if" or "Frightens sincerity John," which are as uninterpretable as they are ungrammatical. Sentence (21) even differs from "John frighted someone," which is ungrammatical even though we can venture a good guess at what someone would have meant by it. No refutation is called for, and any refutation that was attempted would contain a grammatical error.

Putting all of the evidence together, I conclude that the oddness of (21) is due not to ungrammaticality or meaninglessness, but rather to the fact that what it means is something every competent speaker actually or potentially knows to be a category mistake. Even its simple negation, "John does not frighten sincerity," is curious, because we regard it as an

understatement, leaving the hearer to wonder why a stronger statement like (22) or (23) was not used instead. Anomalousness is not meaninglessness or ungrammaticality, but rather a product of the meaning of a compound as it interacts with our knowledge of the world.

The recursive neo-Gricean analysis still needs to be relativized to a language. Chapter 11 will show how to do this without saying that what words mean in English is determined by what users of English mean, which would be circular, given that users of English are defined as those who use words to mean what they mean in English. We also need to acknowledge that word meaning in some languages is not dependent on convention at all.

11

Living Languages

Word meaning is relative to a language. The neo-Gricean thesis that word meaning is recursive, conventional, cogitative speaker meaning holds only for living languages like English. Related theses hold for dead languages and idiolects. Word meaning is determined by stipulation rather than by convention in artificial languages (at least initially), by individual custom in idiolects, and by prior practice in dead languages. What we can say in general is that word meaning is "*established*" speaker meaning. This chapter distinguishes the different ways in which speaker meaning may be established by examining different types of languages. We will focus on the convention-dependence of living natural languages and on the linguistic lineages that their evolving conventions create. The self-perpetuation of conventions coupled with linguistic diversity has generated thousands of genetically related natural languages in much the same way that reproduction and individual variation have generated millions of genetically related biological species. Natural languages are discovered when a community is found that has a previously unknown set of linguistic conventions. What words mean now in a living natural language is determined by the current lexical and constructive conventions of those speakers whose conventions have evolved from the conventions of prior users of the language. This will enable us to avoid the apparent circularity inherent in saying that English is defined by the conventions of English speakers.

§11.1 LINGUISTIC RELATIVITY

The need to exclude ungrammatical compounds (§10.7) forces us to relativize the neo-Gricean analysis to a particular language, for what is grammatical varies from one language to another. There is an even

deeper reason for relativization. Conventions vary from one community to another. It is conventional in America to drive on the right, but not in Britain. Just as it is not conventional for *people in general* to drive on the right, so it is not conventional for people in general to use "red" to mean "red." It is conventional for Americans to use the word "red" in this way, but not for Russians. We must replace *people* with a more specific term in the neo-Gricean analysis. One possibility is *users of e*. The fact that Russians do not conventionally use "red" to mean "red" would then be irrelevant, since most Russians do not use the word "red" at all. And the few who do, use it to mean "red." Unfortunately, infinitely many expressions in any natural language will never be used. Moreover, conventions may vary even among users of expressions that are widely used. As noted earlier, for example, it is conventional in Britain to use the word "bonnet" to mean "hood of a car," but not in America.

We noted in Chapter 7 that word meaning is relative to a language. "Bonnet" means "hood of a car" *in British English*, but not in American English. The formula we really seek to define is:

(1) e means μ in L. (*Linguistic Word Meaning*)

The relativity of word meaning to languages may be obscured by the fact that *e means* μ, without a linguistic reference, is not understood as an incomplete utterance. But that is because it is interpreted with a tacit reference to the object language. When we are talking about English, as we usually are when we use the sentence, *"Vixen" means "female fox"* describes what "vixen" means *in English*. When we are talking about French, *"Renarde" means "female fox"* states what "renarde" means *in French*. The conventional choice of an implicit object language is the natural language to which e belongs (or one of them, should e belong to more than one). But unconventional choices are also possible. If we were to be talking about a secret code in which the word "vixen" also appears, then *"Vixen" means "female fox"* would describe (perhaps falsely) what "vixen" means *in that code*. Statements of the form "e expresses i" *simpliciter* have the same implicit reference to the object language. Recalling §7.3, we have the following generalizations: *e means* μ *iff e means* μ *in the object language*, and *e expresses i iff e expresses i in the object language.* In the same vein, when I use "word meaning" without qualification, I will normally intend "linguistic word meaning" (cf. §7.9 and §11.7).

It is natural to explain what e means *simpliciter* in terms of conventions among *users of the object language*, and to explain what e means *in any language L* in terms of conventions among *users of L*. Thus "bonnet"

means "hood" in British English, but not in American English, because it is conventional for people who use British English to use "bonnet" to mean "hood," but not for people who use American English.

(2) e expresses i in L iff either (i) users of L conventionally use e to directly express i; or (ii) users of L conventionally use $E[x_1, x_2, \ldots, x_n]$ to directly express $I[i(x_1), i(x_2), \ldots, i(x_n)]$, where $e = E[e_1, e_2, \ldots, e_n]$ is a grammatical expression of L, $i = I[i_1, i_2, \ldots, i_n]$, and e_i expresses i_i in L for all i from 1 to n.

This revision of the recursive neo-Gricean analysis (and the equivalent form for meaning) is true in all of the cases we have considered so far. Unfortunately, the claim that such a formula provides an *analysis* of word meaning or expression is problematic due to circularity. We have defined what it is for a speaker to use a language in terms of what the words mean in that language (Definition 7.12). So we cannot now define what words mean in a language in terms of the users of that language. The problem transcends our particular system of definitions. In general, what words mean in English depends on conventions among users of English. How can we identify the users of English except as those who use English words to mean what they mean in English? This is one problem that Grice never saw.[1]

§11.2 CONVENTION DEPENDENCE

The neo-Gricean analysis fails in a trivial way for *dead* (or *extinct*) languages. Completely dead languages are no longer used. What words mean in such languages is determined by how they *used to be* used, not how they *are* being used. The analysis fails more radically for artificial languages that are never used.[2] Basic English, for example, is a form of pidgin English

1 Similar problems have been missed elsewhere. Fodor and Katz (1964: 153–4), for example, asserted that a grammar is primarily a scientific theory of the sentences of a language, and only incidentally a codification of proper usage; the theory is tested against the linguistic intuitions of speakers of the language. But how can we specify who the speakers of the language are without knowing what the language is, and without therefore knowing what the rules of proper usage are?

2 This was observed by Grice (1982: 238–9), who then suggested that "to say what a word means in a language is to say what it is in general optimal for speakers of that language to do with that word." I do not see any sense of "optimal," however, in which this proposal would be both true and informative. Grice (1957: 385) noticed in his original work that automatic signal systems present the very same problem. Red means "Stop!" on a traffic signal, for example, even though no one turned on the red light to mean "Stop!" Cf. Schiffer 1972: 129–30.

devised by C. K. Ogden. But it never became conventional in any community to use Basic English. What words mean in Basic English is determined by Ogden's stipulations, not by conventional usage. Basic English does not qualify as an extinct or even a dead language because it was never alive. It was *stillborn*, we shall say.

The fact that they are being used conventionally is not the full defining feature of living languages, however. The railroad industry once used semaphore signals consisting of rows of lights mounted on posts to indicate how switches on the tracks were set. The semaphore system is a simple language, and was used conventionally. But it was not a living language. A group of people with different native languages who all happen to have learned Ogden's Basic English may well come to use it to communicate among themselves. Even if it became conventional to use Basic English in this way, that would not necessarily make it living.

Why doesn't an artificial language count as a living language even when it is the conventional language of some community? I do not believe the answer is simply that the language had its origin in an act of construction. The fantastic discovery that someone had constructed, say, Basque (a language whose origins are uncertain) should not lead us to conclude that Basque is not a living language. Or suppose that the army began teaching Huron, an extinct Amerindian language, so that it could be used to encrypt secret messages. That would not necessarily bring Huron back to life.

The correct answer, I believe, is that the nature of a stillborn or dead language does not *depend on* the conventions of its users, even when it is used conventionally. The language is fixed by stipulation or past usage. Even if there is a perfect correlation between what words mean in the language and what speakers mean by those words, word meaning does not depend on speaker meaning. To allow for the possibility of change, we were led in §7.1 to represent languages by functions from times to sets of expression-idea pairings. Each value of such a function is a language state. When a language is constructed, what is stipulated is a set of expression-idea pairings. A stillborn language is a *constant* function, whose value at all times is the stipulated set of expression-idea pairings. A living language is a *variable* function, whose value at any given time is the set of expression-idea pairings conventionally used by the language community at that time. The value of the function changes as a consequence of changes in the conventions.

The difference I am pointing to can be illustrated in a slightly different domain by comparing Morse code to the pronunciation rules of a natural

language. The code invented by Samuel Morse is a set of rules pairing sequences of dots and dashes (or long and short pulses) with sequences of letters. Even though it is used conventionally in a number of communities, the correspondence constituting the code is fixed. The correspondence is an external standard, consciously learned and consulted in the process of coding and decoding messages. Pronunciation rules, by contrast, and the orthographic rules that are their inverse vary considerably as conventions vary. Pronunciation rules effect a correspondence between sequences of letters and sequences of speech sounds. The correspondence was presumably stipulated at some point in the distant past, and is taught to children in school. But the correspondence is entirely dependent on convention. People learn the correspondence from other users, and minor differences can and do accumulate from time to time and from place to place. As a result, there are significant regional and temporal differences in pronunciation, known as "accents." Thus "mark" is pronounced "mock" in New York City but not elsewhere. And the pronunciation of Americans is now distinctly different from that of the British. The alphabetic writing system is a living symbol system. Morse code is not.

If natural languages with writing systems were like Morse English, we would have to add a second recursion clause to the neo-Gricean analysis, which would assign meanings to letter sequences on the basis of the rules pairing speech sounds with letters and the meanings of spoken words. But the orthographic conventions of a natural language are not generative in this way. As we observed in §8.5, the general rules for pairing letter sequences with speech sounds in English would have been satisfied equally well if "although" had been spelled "altho" or "althow." The last two letter sequences are not words of English, however, because their usage did not become conventional. Since they are not words of English, they are not grammatical expressions, and have no meaning in English.

Esperanto is one exception that proves the rule.[3] It is an artificial language, invented by L. L. Zamenhof in 1887. Esperanto clubs arose at the turn of the twentieth century, and international congresses were held. Use of the language began spreading around the world. Despite being artificial, Esperanto has been handed down in some families as a first language for three generations. There is an abundant original literature. Esperanto has even begun to develop independently. Colloquial Esperanto differs from Zamenhof's Esperanto. A standard pronunciation

3 Sources: Forster 1982; Large 1985; Anttila 1989: 176; Jordan 1992; Janton 1993: Chapter 4; Wells 1994.

has evolved. The lexicon has expanded dramatically, and neologisms have displaced archaisms. There have been grammatical innovations, such as the use of affixes as independent stems. Scholarly debates have erupted among linguists about factual questions of morphology and syntax, and among teachers about pedagogy. The point might be debated, but I do not think that the evolving Esperanto counts as a natural language: it was invented rather than discovered, and Zamenhof's stipulations are still very much alive in the minds of most users. But Esperanto seems clearly to be a living language.

A more important and complex example is provided by American Sign Language (ASL), which is used by about 500,000 deaf people in the United States and Canada.[4] In the mid-eighteenth century, the Abbé de l'Epée constructed a system of signed French to teach the students in his pioneering school for the deaf. His system was taken to America by Thomas Gallaudet and Laurent Clerc, who established the first American school for the deaf in 1817. Others spread the system around the world, creating cohesive language communities. Because the communities were isolated, l'Epée's artificial language evolved differently in different places. In the French deaf community, it interacted with the preexisting Parisian sign language (or languages) and evolved into French Sign Language (FSL), which is grammatically very distinct from French. In America, the Gallaudet–Clerc system came into contact with the native American sign language system and evolved into ASL. In a result common among pidgins and Creoles, the lexicon of l'Epée's system was used with the basic grammar of the local sign language. Both FSL and ASL are now learned as first languages by large numbers of the congenitally deaf. ASL is more closely related to FSL than to other national sign languages, but the differences are substantial, and users cannot understand each other. ASL is completely unrelated to English or to Signed English, a system for representing English in signs.

We need, then, to augment the definition of living languages as conventional languages with a clause stating the dependence of the states of L on the conventions of the community using L. Just such a clause is provided by the recursive neo-Gricean analysis (§10.3, §10.7). We also need to recognize that the group of users of a natural language is temporally

4 Sources: Woodward 1978; Deuchar 1984: 2–3; Wilbur 1987: 228–9, 251; Fischer 1993: 20–21; Radutsky 1993: 242–8; Schein & Stewart 1996: 18; *Encyclopedia Britannica Micropedia* 1997: 796.

extended, changing over time. If U is the group, let U(t) be the set consisting of all members of U at t.[5] With these two changes, our definition is complete.

11.1 **Definition:** *L is a living language at t iff some group U is such that at t: (i) L(t) is used conventionally by members of U(t) to communicate with one another; (ii) L(t) depends on the conventions in U(t), consisting of all modes of expression ⟨e,i⟩ such that it is conventional for members of U(t) either to use e to directly express i, or to use* $E[x_1, x_2, \ldots, x_n]$ *to directly express* $I[i(x_1), i(x_2), \ldots, i(x_n)]$ *where* $e = E[e_1, e_2, \ldots, e_n]$ *is a grammatical expression of L(t),* $i = I[i_1, i_2, \ldots, i_n]$*, and* e_i *expresses* i_i *in L(t) for all i from 1 to n.*

Clause (i) implies that L(t) changes from λ_i to λ_f during Δt iff the members of U change from using λ_i to using λ_f during Δt. Clause (ii) has a similar implication.

Two different sorts of conventions are mentioned in Definition 11.1 as definitive of living languages. Clause (i) requires a *global* convention to use the whole language (state) L(t).[6] Clause (ii) requires *local* conventions to use and interpret each word, idiom, and construction in L(t). The two sorts of conventions go together naturally, of course, but the clauses are logically independent. As we observed in §7.8, an individual who uses a standard natural language cannot possibly use every word in that language, not even every simple word. Hence clause (i) does not entail that for each and every simple word in L(t), there is a convention to use it in a certain way, which is required by (ii). Clause (i) would allow there to be some words in a living language that were never used themselves, and were not even composed of words that have been used. Clause (ii) makes this impossible. On the other hand, satisfaction of condition (ii) ensures only that it is conventional in U(t) to conform to L(t), which is less than what (i) requires (see §7.8). Note finally that it is possible for a person to use English today without using today's English. Definition 11.1 implies that today's English depends on a group's using English today.

5 The group is "modally extended," too: the membership of the group might have been different, and would have been if certain conditions had been different. But we will focus on the actual world.

6 Since Lewis (1969; 1975) did not define what it is for an individual to use a language, or take the notion to be primitive, he was unable to use such global conventions. Cf. Schiffer 1982: 123.

Definition 11.1 entails that the recursive neo-Gricean analysis holds for living languages like English.

11.2 **Theorem:** *If L is a living language, then e expresses i in L iff (i) the people on whom L depends use e conventionally to directly express i, or (ii) they use* $E[x_1, x_2, \ldots, x_n]$ *conventionally to directly express* $I[i(x_1), i(x_2), \ldots, i(x_n)]$*, where* $e = E[e_1, e_2, \ldots, e_n]$ *is a grammatical expression of L,* $i = I[i_1, i_2, \ldots, i_n]$*, and* e_i *expresses* i_i *in L for all i from 1 to n.*

Similar principles relate "e means μ in L" to "By e S means μ," "e refers to R in L" to "S uses e to refer to R," and so on. Basically, a language is living only if the recursive neo-Gricean analysis holds for it relative to a certain group of people, whose semantic-act conventions define the language.

Theorem 11.2 defines word expression for living languages. Given the connection between meaning and expression (Definition 7.4), it tells us what it is for words in a living language to have meaning. Theorem 11.2 entails, for example, that "vixen" means "female fox" in English because the language users on whom English depends use "vixen" to mean "female fox." This is not circular, even though the people on whom English depends are users of English, and users of English are those who use "vixen" to mean "female fox." Theorem 11.2 is noncircular because of the convention-dependence of living languages. The people on whom English depends are identified in terms of their historical relations to prior users – as people whose linguistic conventions descended from those of prior speakers of English. Since the prior users can be identified without knowing that they have used "vixen" to mean "female fox," no circularity arises.

Conventions are regularities that perpetuate themselves in a certain way. They obtain at a given time t only if they have perpetuated themselves in that way recently. Hence clauses (i) and (ii) of Definition 11.1 imply that members of U at one time followed the conventions of L because members of U before them did. Indeed, Definition 11.1 implies that U is the temporally extended community picked out by the self-perpetuating conventions of L.

Since the regularities need not be perfect, it is possible for conventions to change while perpetuating themselves. Only sudden, radical changes

are ruled out.[7] For example, if all Americans voted tomorrow to switch from English to Japanese, and did so, American English would not turn into Japanese. Definition 11.1 yields this result, because there would be some period of time during the transition during which neither the regularities of Japanese nor the regularities of English were perpetuating themselves in the right way. Americans would no longer consider the fact that Americans had previously spoken English a good reason to keep on speaking English. And for a while children would not be able to pick up Japanese naturally in the home. Associations and habits would be disrupted. Hence Americans would speak Japanese not because previous Americans had done so, or out of habit, but because of the vote.

Given the self-perpetuating character of conventions, the group U on which L depends is a *lineage*, consisting of the *original* speakers of the language and their linguistic *descendants*. As a consequence, every subsequent state of a natural language results indirectly and in part from the practices of the original group of speakers. When we consider how a natural language would have evolved under different conditions, we must consider what the line of descendants would have been under those conditions. The line of linguistic descendants will be confined to biological descendants only when the group remains extremely isolated. Normally, the biological descendants teach their language to immigrants, who pass the new language on to their offspring, in a repeating cycle. Conversely, emigration removes many biological descendants from the linguistic lineage.

A new living language is discovered and named when a group of people is found with a previously unknown set of semantic conventions (§11.6). The referent of the name is fixed by a description like "the language this community uses." The community's own name for its language may be fixed by a description like "the language we speak." The conventions defining the modes of expression of the language are those that evolve from the conventions of the original group in the ways described in §9.3. The people on whom the language depends are those whose conventions descend from those of the original group. The people on whom L depended in earlier times were those whose semantic-act conventions

7 "There are just two principles of making inferences about time past. Either great differences, or wide geographical spread between languages requires time" (Anttila 1989: 386). Cf. Woodward 1978: 337; Nurse & Spear 1985: 9. See also Halle 1962: 346: "Linguistic change is normally subject to the constraint that it must not result in the destruction of mutual intelligibility between the innovators – i.e., the carriers of the change – and the rest of the speech community."

evolved into those of the group relative to whom the reference of "L" was fixed. Since "L" and "the people on whom L depends" are not defined in terms of particular modes of expression, there is no circularity in the claim that what an expression means in L is determined by the meaning conventions found in the group on whose usage L depends.

L is a living language only if there is *some* line of users on which it is dependent. This group need not be, and usually is not, the *entire* set of users. Consider a standard science fiction scenario. A distant planet called Zarn is much like Earth. Parallel evolution has given rise to organisms like ourselves, who happen to be speaking English. Such a coincidence would, of course, be fantastically unlikely.[8] The observation I wish to make, however, is that even if there were such extraterrestrial speakers, the future development of English would depend solely on what happens to the conventions of English speakers on Earth. If the usage of Zarnians should begin to diverge from that of earthlings, which remained static, English would remain static. If the divergence became sufficiently great, the Zarnians would no longer be speaking English. The same point can be made without leaving Earth, of course. (Suppose a native speaker of Chuvash sat down to construct an ideal language, ended up with English, and convinced all the natives of the Autonomous Republic of Chuvash to adopt it.)

One familiar and often large group of users on which a language does not depend are those just learning the language, or using it awkwardly and infrequently as a second language. Millions of adults around the world are taught English as a second language. The development of English does not depend on their usage, at least until they become highly proficient. If teaching methods were completely effective, changes in the usage of such a group would be perfectly correlated with changes in English. But the direction of dependence would be the reverse of that mentioned in Definition 11.1.

The most common reason that a group of speakers exerts no influence on the development of a natural language is isolation – geographic, social, or political – from the main line of speakers. When the British colonists settled America, they spoke British English. Subsequent changes in their usage altered English, but had no direct effect on British English. There

8 Linguists are justified in adopting, as an inductive methodological rule, the principle that "[i]f two or more languages share a feature which is unlikely to have occurred spontaneously in each of them, this feature must have arisen once only, when these languages were one and the same" (Anttila 1989: 302. Cf. Nurse and Spear 1985: 7).

was presumably some indirect effect, since speakers in Britain were bound to pick up some Americanisms, especially when large numbers of subjects returned to Britain. Similarly, the Boers who migrated to South Africa spoke Dutch. Their usage changed, in part owing to contact with native African languages. This gave rise to Afrikaans, without affecting Dutch (except to the extent that the main line of Dutch speakers was influenced by the Afrikaners).[9]

We have noted that the primacy of the spoken component of natural languages has led some to conclude that writing is not even a part of language (§7.5). A related fallacy – somewhat weaker, and considerably more plausible than the first – is that written expressions are not part of the language they are associated with. We need to address this, because Definition 11.1 allows written expressions to be meaningful in living languages. Once speakers of a natural language become literate, Definition 11.1 rules that the written expressions are part of the language. For L(t) contains all pairs ⟨e,i⟩ such that members of U(t) conventionally use e to express i. Despite the apparent absurdity of claiming that written English is not English, arguments have been offered to that effect.

> A language is the same no matter what system of writing may be used to record it, just as a person is the same no matter how you take his picture. The Japanese have three systems of writing and are developing a fourth. When the Turks, in 1928, adopted the Latin alphabet in place of the Arabic, they went on talking in just the same way as before. In order to study writing, we must know something about language, but the reverse is not true. (Bloomfield 1933: 21)[10]

First, writing is not merely a recording of speech, as if a stenographer were making a record of our testimony. We express ourselves in writing as genuinely as we do in speech. For literate people, the mode of expression is normally a matter of choice. Sometimes we prefer spoken expression, sometimes written. The choice of medium predicts other linguistic features, such as the relative frequency of contractions, learned locutions, hesitation forms, broken sentences, and the average length of sentences.

9 One scholar has argued that "Dutch and Flemish . . . are so similar that they should be regarded as the same language although they are spoken in different political entities" (Muller 1964: 2); see also Vogelin et al. 1988: 663. But the evolution of Flemish and Dutch depends on distinct user-groups, and consequently could evolve in radically different ways. So the languages cannot be numerically identical. Neither, moreover, is identical with Dutch-Flemish (Netherlandic), the more inclusive language of which Dutch and Flemish are distinct dialects.

10 Cf. Algeo 1973: 17; Landau 1984: 207–8; and Anttila 1989: 31.

Second, it is undeniable that Turkish survived the change of alphabet that occurred in 1928. The Turks used Turkish both before and after 1928. But it is equally undeniable that the Turkish language changed in 1928. For there was a sweeping change in Turkish orthography. While Turks went on talking in the same way, they did not go on writing in the same way. Numerical and substantial identity through qualitative change is a universal phenomenon. The fact that our body survives a nose job does not imply that our nose is not part of our body. I can see no scientific basis whatever for saying that trivial differences of pronunciation produce different dialects, while radical changes in orthography have no effect on the language. Note that some of the most widely known differences between British English and American English concern spelling: "colour" versus "color" and the like. German differs from English in capitalizing common nouns wherever they occur. In the same vein, when our son was deciding whether to take Japanese or French in high school, he was advised that Japanese is harder to learn because of the multiple writing systems. It is true that spoken English is one thing and written English another, and that the former has more influence on the latter than the latter has on the former. My point is that there is as much reason to classify written English as part of a natural language as there is to so classify spoken English.

§11.4 BOUNDARIES

All living languages had a beginning, and undoubtedly all will have an end. The beginning of a language is the time at which people begin to speak it. The end is the time at which people stop using the language, because it has either evolved into or been displaced by another language. Old English began in the fifth century, and evolved into Middle English in the twelfth century. Cornish became extinct at the end of the eighteenth century, displaced by English. The beginning of a language is represented by the first time at which the language function is defined. The end is represented by the last time at which the function is defined. Before a language originates, and after it terminates, it has no states to be dependent on the conventions of any group. Words do not mean anything in the language at those times. That is why clauses (i) and (ii) of Definition 11.1 need not hold for all times. Present-tense statements about a language normally refer to its current state. Thus *"Jument" means "female horse" in French* is true because that is what "jument" means today, although earlier it meant "pack horse." In the case of dead languages, present-tense

meaning claims may refer to a standard period. Thus *"Testa" means "earthenware bowl" rather than "head" in Latin* would ordinarily be understood as a true statement about Ciceronian or Classical Latin, even though "testa" had acquired the meaning "head" in Late Latin, and despite the fact that the language is dead today.[11]

Nothing in the concept of a natural language tells us how similar the initial and final states of a living language must be, just as nothing in the concept tells us how similar different dialects must be. For example, the term "German" is ambiguous. In a narrow sense, it refers to modern German, which arose around 1500. In a broad sense, it refers to the historical progression beginning with Old German around A.D. 750, evolving into Middle German around 1100 and then Modern German around 1500. In both senses, German is a living language and satisfies Definition 11.1. We could, if we liked, distinguish Early Modern German (before 1800, say) from Late Modern German. There may be no good theoretical or practical purpose for making such a distinction, of course. But there would also be no ground for denying that either Late Modern German or Early Modern German is a living natural language. German in the narrow sense is but a *phase* or *stage* of German in the broad sense, and Early Modern German, thus defined, is a stage of both. L(t) is a stage of L′(t) iff L(t) = L′(t) for all times at which L(t) is defined, assuming that both functions are defined on continuous intervals. Whereas a "state" of a language is a value of the language function, a "stage" is a function whose domain is limited to part of the time interval over which the whole language is defined, and whose values are the same over that subinterval.

Early Modern German and Late Modern German are different languages in the sense that they are not numerically identical. The phrase "different languages," however, normally connotes unrelatedness or significant qualitative dissimilarity, both of which are inapplicable in this case.[12] Thus Simpson (1994: 1894) cited the fact that "'on the ground' there is no sharp break between so-called 'Dutch' dialects and so-called 'German' dialects" as proving that Dutch and German are not "distinct" or "different." What the cited fact proves is that the two languages are not unrelated or radically different. The fact that some dialects of Dutch are

11 Löfstedt 1959: 26; *Cassell's New Latin Dictionary* 1959: v, 601.

12 Consider an analogy, from continental drift. Los Angeles moved northwest along the San Andreas fault between 1800 and today. It also moved northwest along the fault between 1600 and today. Both were tectonic movements, and were clearly not identical. Yet the claim that they are different movements seems incorrect. For that would normally imply that the movements were unrelated or dissimilar, which they clearly were not.

not dialects of German, and vice versa, proves that Dutch and German are nonidentical, and both facts are compatible with the practical continuity of their dialects. One might just as well try to prove that blue and green are not different colors because there is no sharp break between the bluish shades and the greenish shades.

Establishing a unit of linguistic difference is important for some purposes, such as estimating the total amount of linguistic variation found in the world. This is what I believe people are really interested in when they ask how many languages there are. A unit is also needed in order for historical linguists to estimate the time at which two languages diverged on the basis of the degree to which they differ today.[13] Defining a unit of linguistic difference is an exceptionally difficult and perhaps impossible task, given the complex and multidimensional nature of natural languages.[14]

One attractive measure is provided by the notion of "mutual intelligibility." Languages that are mutually intelligible differ less, by this criterion, than those that are not. As a measure of overall similarity, this criterion seems to be defective, since one linguistically minor difference sometimes produces mutual unintelligibility (cf. Halle 1962: 342–3, on pig Latin). And whether or not one dialect is intelligible to a speaker of another dialect depends on nonlinguistic factors such as the linguistic aptitude of the hearer, the extent of her exposure to the dialect, and so on. Mutual intelligibility cannot be used, moreover, to provide an absolute distinction between a dialect and a language, or between a language stage and a language. For dialect (or stage) A may be mutually intelligible with B, which is in turn mutually intelligible with C, even though A is not mutually intelligible with C (Cf. Ziff 1960: 3–4). The dialects of German and Dutch present many such cases. For our purposes, the most important point is this: the fact that dialects (or stages) A and B are mutually intelligible does not entail that A and B are not living languages. A unit of linguistic difference is not needed for the purpose of saying what a living language is.

It is customary to distinguish two principal dialects of German, High and Low. The latter is spoken at the lower elevations of northern Germany, the former at the higher elevations to the south. High German resulted from a "sound shift."[15] Low German "p" became High German "f," as

13 Vogelin et al. 1988: 605.

14 Cf. Muller 1964: 1–3; Woodward 1978: 337; Nurse and Spear 1985: 9; Vogelin et al. 1988: 591; Anttila 1989: 386.

15 Muller 1964: 16–18; Anttila 1989: 290; Vogelin et al. 1988: 665.

in *Dorp/Dorf* ("village"). Low German "t" became High German "s," as in *dat/das* ("that"). Low German "k" became High German "ch," as in *maken/machen* ("make") and *ik/ich* ("I"). There is no precise geographic boundary north of which all features of Low German are present, and south of which all features of High German occur. What we find instead are "isoglosses" for each of the features, which gives rise to the notion of "dialect geography." The isogloss for *Dat/das* crosses the Rhine farthest south, near Coblenz. North of this line, Germans say "Dat," below it, "Das." The isogloss for *Dorp/Dorf* crosses the Rhine farther north, near Bonn. "Dorp" is used north of the isogloss, "Dorf" south of it. The isogloss for *maken/machen* is farther north, near Cologne, and that for *ik/ich* farther north still. As these facts make clear, there are more than two dialects of German. High German and Low German represent two points on an almost continuous series.

It seems likely that an analogous distribution occurred over time at any one location where High German is now spoken. That is, "k" probably changed to "ch" first, then "p" to "f," and finally "t" to "s." It is also likely that each of these shifts occurred word by word, with "ik" changing to "ich" before "maken" changed to "machen." If so, then just as there is no north-south boundary at which we can say "Here and only here a different dialect begins," so there was no date in the past at which we could say "Then, and only then, a different language began."

The boundaries of a language are thus vague for two reasons. First, the dating of nearly all linguistic changes is inherently imprecise. It is not possible, for example, to give a precise date at which the High German sound shift began, or a date at which Germans near Coblenz changed from "Dat" to "Das." No matter how much evidence we have, it is not possible to specify a precise moment before which a certain action was conventional, and after which it was not. For this reason, if a language is defined indirectly in terms of a group of users and linguistic features, rather than directly by assigning expression-idea pairings to times, then it will be impossible to specify exactly when the language arose. For example, Middle High German is customarily defined as beginning when another sound shift, involving vowels, had advanced to a certain point among speakers of German.[16] This occurred around A.D. 1100, but the figure 1100 is an arbitrary one. No amount of evidence will enable us to say down to the year, not to mention the second, exactly when Middle High German arose.

16 Vogelin et al. 1988: 666.

Second, there is no inherent reason why one set of linguistic features rather than another should be taken as defining a language. Different features could have been selected, resulting in the definition of a language with a different date of origin. If a less advanced point of the shift were selected, for instance, a stage of German would be defined whose (vague) date of origin was earlier than that of Middle High German. But it would be as much a language as either German or Middle High German. Indeed, it would be a stage of German, and Middle High German would be a stage of it.

Finally, suppose that for centuries a group U spoke a common language. After t, two subgroups, U_1 and U_2, became separated and developed different languages. There are several equally correct ways to represent this situation. (1) U spoke A before t, which evolved into two new languages, B and C. This implies that at some point U_1 and U_2 ceased using A. (2) U spoke D before t, which developed two dialects, E and F. This implies that U_1 and U_2 never stopped speaking D, which contained E and F after t. (3) U spoke G before t, and U_1 continued to speak G. U_2 developed a new language, H. (4) U spoke I before t, and U_2 continued to speak I. U_1 developed a new language, J. There may of course be reasons why one of these representations is more intuitive or useful than another. For example, if U_2 was a relatively small and socially unimportant part of U, and if conventions changed slowly in U_1 and quickly in U_2, then (3) is the most natural representation. On the other hand, if U_1 and U_2 were roughly comparable groups, which remained in communication with each other, and if the conventions in U_1 and U_2 changed relatively little, then (2) is the natural representation. If U_1 and U_2 were roughly comparable, but after a while could no longer communicate, then (1) is the most natural. The fact that one representation is the most intuitive or useful does not, however, entail that the other representations are incorrect. At least six different functions from times to expression-idea pairings can be distinguished in this situation.[17] All represent living languages.

§11.5 LANGUAGE DEATH

A living language becomes *dead* (or *extinct*) when it either ceases to be conventionally used, or ceases to be convention-dependent. A language

17 I say at least six rather than ten, because given only what I have said, language B might be identical with E and J, while C is identical with F and H. In that case, A would be a stage of D, G, and I.

is *completely dead* when it ceases to have any users at all. Thus Latin is a dead language, although, unlike Etruscan, it is not completely dead. Viewing death as a process rather than a state, we can say that Latin is in the later stages of a protracted death.[18] Colloquial Latin, used in daily life, began to diverge from the formal Latin used in literature and oratory before the classical period of Cicero (106–43 B.C.) and Vergil (70–19 B.C.). Between the third and eighth centuries A.D., colloquial Latin evolved into the modern romance languages. Literary Latin continued in a standardized form throughout the Middle Ages. Learned as a second language, it was the language of religion and scholarship. Latin was used for science until the sixteenth century, when it began to be replaced by modern national languages. Most scholars continued to learn Latin in school until World War II, when its popularity declined more sharply. The requirement that Latin be used in the liturgy of the Catholic Church was dropped even more recently. Latin is still used conventionally by many in the church, but I do not believe that it is changing as a result of conventional usage. Latin is no longer a living language, but it will not be completely dead until it is no longer used at all.

This "latinate" or "bottom-to-top" pattern of language death, where the language is lost first in the informal registers used in contexts of domestic intimacy and later in the formal registers used in literature and ritual, is common. "Top-down" language death also occurs, however, particularly when a language dies under the pressure of an imperial or national language, but is retained for a while as the language of solidarity. Thus dialects of Nahuatl, an Indian language of Central America descended from Aztec, have died in many communities as speakers gradually shifted to Spanish. At first, bilinguals used Spanish exclusively for trade. Later, Spanish became the public language, first in religion, then in government and schools. Finally, Nahuatl lost even its solidarity function and died completely.

Dying languages undergo the same sorts of changes that healthy languages experience, especially those in contact situations.[19] The number

18 Sources: Löfstedt 1959; L. Campbell 1994; Mesthrie 1994; Posner 1997; Baldi 1994; Dressler 1996. *Webster's Encyclopedic Unabridged Dictionary* defines a dead language as one "no longer used as a sole means of oral communication among a people." By this "sole use" criterion, Latin died between A.D. 600 and 800. But by the same criterion, Romansh would not count as a living language if, by a quite realistic chance, its native speakers happened to use German, French, or Italian as well. L. Campbell (1994) and Mesthrie (1994) define a language as dead "when it no longer has any speakers." By this criterion, Latin is not dead, and Etruscan would not be dead either if scholars began speaking it.

19 Dorian 1981; L. Campbell 1994: 1962; Mesthrie 1994: 1991–2; Dressler 1996.

of tense, aspect, and mood distinctions is often reduced, and case systems decay. Infrequently used words are lost and often replaced by words from the dominant language. They may even be revived, as Hebrew illustrates. Death is sometimes practically instantaneous, however, as in the case of Tasmanian.[20]

Finally, perfectly healthy languages may become dead when they evolve into new languages. Cornish, the Celtic language of Cornwall, became extinct around 1800, when children of Cornish speakers began learning English rather than Cornish and the last adult speaker of Cornish died. Children of Old English speakers, by contrast, continued to learn the language of their parents. There was no sharp break as Old English turned into Middle English, and no one can be positively identified as the last speaker of Old English. Old English is a superseded stage of a continuing lineage, whereas Cornish was the end of a terminated lineage.

§11.6 NATURAL AND ARTIFICIAL LANGUAGES

Nearly all living languages are natural languages. Whereas artificial languages are invented, natural languages are discovered. When a living natural language is first found, what we discover is a group of people using a common language. We realize that they are using a common language before we know what the observed speech sounds mean in it. We determine what the expressions mean by observing actual usage, and figuring out what it is conventional for speakers to express when they use the words and constructions of the language.

Once we realize that a common language is being used, we can name it. While the reference of the language name is fixed in terms of the group of individuals observed to be using it, the term functions as a rigid designator of the abstract system constituting the language. Suppose we find a group of people named the Unias speaking an unfamiliar language, which we name Unia. It is then true by definition that Unia is the language of the Unias. Nevertheless, the Unias could have spoken a different language, and different people could speak Unia.

While a language need not be used by any *particular* people, it does not count as a natural language unless it is used by *some* people. Natural

20 The Tasmanians were a population of Australian Aborigines, isolated on a large island to the southeast of Australia, who spoke languages with no known relatives. Numbering about 4,000 in 1803, when Europeans first settled, all died of disease or were killed. The last speaker of Tasmanian died in 1876. Sources: Campbell 1994: 1960; *Encyclopedia Britannica* 1997: 572.

languages are found in the natural world, and are not merely objects of human conception. A natural language, moreover, is generally *used "naturally"* in the sense that some people use it spontaneously, effortlessly, unself-consciously, automatically, and habitually. Its use is *"second nature."* Indeed, a natural language is generally used naturally because it is the *native language* of the speakers. Until they become dead or extinct, natural languages are passed on from one generation to the next during the first few years of life as part of the normal maturation process. Because all languages must be learned, and different people learn different languages depending on their native environments, even natural languages are not *innate*. Only the capacity to learn a language is innate.

Dictionaries define natural languages as those that are used natively by some group of speakers.[21] This is satisfied by typical natural languages, but not by all. Counterexamples include the numerous *pidgin languages*, such as Hawaiian Pidgin English, Chinook Jargon, or the original Lingua Franca. Such languages evolved when peoples with different native tongues had to communicate, usually for commercial purposes. Only in a few cases have pidgins become native languages, a process called "creolization." The resulting *Creoles*, such as Hawaiian Creole English, appear to be more complex and powerful than the pidgins they resulted from (Rickford 1992: 226; Bickerton 1981: Chapter 1). But it does not seem correct to say that only the Creole is a natural language. Furthermore, at least one artificial language is used natively by a group of speakers, and that is Esperanto (Anttila 1989: 176).

A natural language can be better defined, I submit, as a nonartificial language that is living or dead.

11.3 **Definition:** *L is a natural language iff L is a nonartificial language that is or was living.*

An artificial language is one that was stipulated by some individual or group. It consists of all the expression-idea pairings linked by the stipulations of that individual or group, or evolved linguistically from a stipulated language. When more than one individual is involved, the stipulation constitutes an *agreement* (cf. §9.1).

11.4 **Definition:** *L is an artificial language iff either (i) L depends on the stipulations rather than customary usage of some individual or group U,*

21 See, for example, *Webster's Encyclopedic Unabridged Dictionary* and *Webster's Third New International Dictionary*.

consisting of all pairs $\langle e,i \rangle$ *such that U stipulated either that e means* μ_i *in L, where "*μ_i*" expresses i, or that* $E[x_1, x_2, \ldots, x_n]$ *means* $E'[x_1, x_2, \ldots, x_n]$ *in L, where* $e = E[e_1, e_2, \ldots, e_n]$ *is a grammatical expression of L, and* $E'[e_1, e_2, \ldots, e_n]$ *expresses i; or (ii) L descended without much evolution from a language satisfying (i).*

Meaning in artificial languages is typically not determined by conventions. There is no requirement that L be used by anyone, or that it vary over time. The case of Esperanto shows that artificial languages might come alive, and evolve. If the change should become so extensive that few if any of the originally stipulated modes of expression were part of the language, then I believe it would not longer qualify as an artificial language. A natural language would have been born. ASL and FSL surely qualify as natural languages.

We noted in §9.6 that a use of language is "correct" provided that the speaker uses modes of expression from the language that he or she is trying to use. Hence conventional usage is the proper standard of correctness for English, since English is a natural language. Whether a mode of expression is in a stillborn artificial language, by contrast, is determined by stipulation. So an attempt to use the language is correct only if it conforms to the stipulations defining the language.[22] If an artificial language comes alive, there are two standards of correct usage: the stipulations defining the original language, and the evolving conventions.

One argument that Definition 11.3 is too narrow starts from the premise that *innate* languages are possible. For example, Blackburn asks us to consider "the possibility of groups whose linguistic regularities are 'wired in,' so that they naturally take certain sounds to have certain meanings; the regularity would not have the conventional status, but it might be arbitrary to deny that they have a language and genuinely attribute meaning to the features" (1984: 122).[23] Suppose it is also "wired in" that whenever a speaker thinks certain thoughts (or wants to provide an indication that he is thinking them), he automatically makes certain sounds. Even if it is granted that this counts as using a language, the possibility would at most show that Definition 11.3 does not state *logically* necessary conditions for a language to be a natural language. Definition 11.3 may still correctly pick out the set of all actual natural languages. But it is not clear that the behavior Blackburn describes would count as *language*, especially

22 Cf. Gilbert (1996: 78–86, 107–12), who notes an analogy between the "ought" of agreement and that of convention.

23 See also Hockett 1966: 13 and Peacocke 1976: 169.

since it is involuntary. Moreover, the imagined speakers are not expressing their thoughts, and do not mean anything by their words. Their behavior is only an *evidential* (or natural) expression of their thoughts, rather like blushing (cf. Chapter 3). We can certainly refer to the set of expression-idea pairings picked out by their behavior as *a* language. Indeed, that language may have the characteristic syntactic and semantic structure of natural languages. But it would be hard to describe it as a natural language, given that the behavior of its "speakers" is only superficially like the actual behavior of speakers of paradigm natural languages like English. It would seem even more strained to call it a "living" language. In any event, since such a language is at best a far-fetched possibility, I am not inclined to give it much consideration.

Another feature of the dictionary definition is inappropriate for our purposes. Natural languages are typically *spoken*. And in those that are written as well, the written language is in many ways secondary. Nevertheless, being spoken is not an *essential* feature of a natural language. In the event that it were no longer spoken, English would survive if it continued to be written. Like Latin, it might begin to die. But it would not become extinct. More decisively, ASL, FSL, and the other national sign languages surely qualify as natural languages.[24] By contrast, a language that is never used by any group of people ipso facto fails to count as a natural language.

Many other linguistically important features common to all natural languages are logically inessential.[25] For example, all natural languages exhibit *duality of patterning*: the thousands of words in the language are arrangements of a relatively small number of basic units, which themselves have no meaning in the language. Even if all natural languages have this feature as a result of the nature of the human mind and body, duality is not logically entailed by the claim that a language is a natural language. If we were to discover a group of people using natively a language in which the smallest syntactic unit is meaningful, there would be an exception to the rule. Similar comments apply to the other well-known formal features common to all natural languages, such as *discreteness* and *productivity*.

I hasten to add that since our primary concern is with the relation between word meaning and speaker meaning, nothing that we discuss here would be affected by adding to Definition 11.3 a clause specifying

24 See, for example, Siple 1978; Deuchar 1984; Wilbur 1987; Schein and Stewart 1996.

25 See, e.g., Chomsky 1962: 223–4; Katz 1966: 107; and Anttila 1989: 27. For a more inclusive list, see Asher (1994), who discusses sixteen "design features" of language proposed by Hockett (1958: 574; 1966), and R. Harris (1980: 22–9), who criticizes the design-feature approach.

that L be spoken or have features such as vocality, duality, discreteness, and productivity. And since all natural languages in fact have these features, the same group of languages would count as natural if such a clause were added.

§11.7 IDIOLECTS

Somewhere between artificial and natural languages are *idiolects*. As Grice (1968: 226) observed, locution (3) differs from both "S means μ by e" and "e means μ," while having affinities with both. We will call it *idiolectic word meaning*, or "idiolect meaning" for short.

(3) e means μ for S. (*Idiolectic Word Meaning*)

Idiolect meaning is a form of word meaning. In (3), the subject of the verb "means" refers to the word, not to the speaker. Nevertheless, idiolect meaning differs from linguistic word meaning: if S regularly uses a code in which "sunshine" means "cocaine," then "sunshine" means "cocaine" for S even though "sunshine" has a different meaning in English. Conversely, suppose S made up a "Code B" in which "sunshine" means "bourbon," but never uses that code. Then "sunshine" means "bourbon" in S's code B, but does not mean "bourbon" for S.[26] Idiolect meaning also differs from cognitive speaker meaning:[27] If S says "My car is running well" ironically, he means that his car is running poorly. But what the expression means for him is just "My car is running well." Idiolect meaning differs even from cogitative speaker meaning in requiring a certain type of regularity over time: If S suffers a slip of the tongue when he says "erotic," he may have meant "erratic" even though "erotic" does not mean "erratic" for S. "Erotic" does not mean "erratic" for S because S does not *customarily* use the word "erotic" to mean "erratic." He did on this occasion, but only by mistake. The sentence "e means μ for S" is equivalent to "S means μ by e" only when the latter is taken dispositionally, in its frequentive sense.

Conventions are *social customs* (or *practices*): regularities in action among many people, which are arbitrary but which serve a common interest and transmit themselves from individual to individual and from generation to generation. *Individual customs* (or practices), by contrast, are *regularities over*

26 Cf. Grice 1968: 234, case 3.

27 That figurative usage may be customary for S is a problem for Grice's (1968: 233) suggestion that idiolect meaning is customary cognitive speaker meaning. Another problem is the productivity of idiolect meaning (see the following discussion).

time in the voluntary actions of one individual, which are arbitrary but useful for that individual and self-perpetuating. An individual custom is a regular way an individual has of doing something, which serves that individual's purposes. The regularity is self-perpetuating in the way that habits are. Moreover, the individual considers that manner of doing things to be the correct way. And the fact that he has acted that way in the past gives him a good reason to continue doing so. Finally, individual customs are arbitrary in the same way that group customs are: there is another regular way of doing things that would serve the same purposes, and would perpetuate itself in the same way.

Whereas linguistic word meaning is established by social custom (or stipulation), idiolect meaning is mainly established by individual custom.[28] This suggests that (3) is equivalent to (4):

(4) e means μ as used by S (customarily).

What (4) says is that S customarily uses e to mean μ. Hence (4) expresses the notion of customary cogitative speaker meaning. The notion of idiolect meaning is much broader, however. Idiolect meaning may be established directly, by customarily using a certain expression to express a certain idea. But as in the group case, idiolect meaning may also be established indirectly, by customarily using the component words and constructions of a complex expression. Indeed, idiolects normally do have construction rules. There is a third way in which e may get a meaning for an individual, moreover. Most people consider certain words to be profane or obscene. Due to strict upbringing, some never use them. Nevertheless, the words have their conventional meaning for them. Indeed, it is because the words have the meaning they do for them that such people refuse to use swear words. But if the words are never used by S, and are not composed of other words used by S, then their meaning for S cannot be established by S's customary usage.[29] Instead, they have a meaning for S because S *takes* them to have an established meaning *in a language that S customarily uses*. As defined in §8.6, to take e to mean μ is to think of e as meaning μ in some way, whether one believes that e means μ or simply hears e as meaning μ. In the group case, customary usage coincides with customary interpretation. In the individual case, the two customs may diverge. This gives us:

11.5 **Definition:** *e means μ_i for S iff e expresses i for S.*

28 Cf. Grice 1968: 231–3 and Blackburn 1984: 88–91.

29 Cf. Grice's (1968: 234) prim Aunt Matilda.

11.6 **Definition:** *e expresses i for S iff S customarily uses or takes e to express i, or E[x_1, x_2, . . . , x_n] to express I[i(x_1), i(x_2), . . . , i(x_n)], where e = E[e_1, e_2, . . . , e_n] is a grammatical expression for S,*[30] *and e_i expresses i(e_i) for S for all i from 1 to n.*

No one, not even the most knowledgeable native speaker, knows the meaning of every word in English. If S has no idea what e means in the language he customarily uses, then e has no meaning in his idiolect. Every speaker, moreover, will be mistaken about the meaning of some words. If S mistakenly believes that e means μ, then e means μ for S, assuming that his belief is reflected in his interpretive practice. That assumption is important. Because idiolectic meaning is established by individual custom, just as linguistic meaning is established by social custom, it is possible for an individual to be mistaken about what a word means for him, just as it is possible for him to be mistaken about what it means in his language.

S's idiolect may be defined as the language in which expressions mean what they mean for S, and may therefore be represented by the function whose value at any given time is the set of all pairs ⟨e,i⟩ such that e expresses i for S at that time.

11.7 **Definition:** *L is S's idiolect iff e means μ in L(t) provided e means μ for S at t.*

It is true by definition that one *conforms* to one's own idiolect, but not that one *uses* it (see §7.8). S uses his own idiolect only if he has some conception of it, and only if he uses words because of what they mean in his idiolect. The typical situation, I believe, is for speakers to use words because of what they mean in a natural language like English. Note that a natural language will seldom, if ever, be the union of the idiolects of all of its speakers. For there will usually be speakers of the natural language whose customary practices are unconventional.

§11.8 CONVENTIONS OF TRUTHFULNESS

According to Lewis (1969: 194), a community C uses L iff there prevails in C a convention of truthfulness in L, sustained by an interest in communication. Let us say that to be *truthful in L* is to use L to tell the truth – to use L to express what one takes to be true. Someone who uses "The

30 We will not attempt to analyze what it is for a complex expression to be grammatical for an individual or in a natural language.

defendant was at the scene of the crime" to express a belief and thereby tell the truth, is being truthful in English. Nearly everyone lies sometimes, of course. And some people may lie whenever it serves their interests. But most people, even those we call liars, tell the truth most of the time. This generalization holds, I presume, within every community in which there is a conventional language. Given that people are generally truthful in any language community C, it follows that L is used conventionally in C if and only if it is conventional for members of C to be truthful in L.

Being truthful *simpliciter* is not a matter of convention.[31] People generally consider truthfulness to be a moral imperative, and often a practical necessity. The fact that others before us have told the truth is seldom a relevant consideration, let alone a good and decisive reason, for being truthful ourselves. The interests served by general honesty could not be served by the alternative regularity, general dishonesty. That is, there is nothing arbitrary about the practice of telling the truth. Nevertheless, the use of a particular language to tell the truth is every bit as conventional as the use of that language. We could just as well use French rather than English to tell the truth, except for the fact that others in our community use English.[32]

The existence of a convention of truthfulness in L is not necessary in order for L to be a living (or natural) language. For the existence of a community in which L is conventionally used to tell the truth is not logically necessary in order for L to be a living language. First, the mere

31 Cf. Fodor 1983: 45. Contrast Barwise & Perry 1983: 18: "Obviously, if truthful assertions were not an important part of our life, and if we did not possess a fairly good ability to recognize them, utterances would not carry information for us, and so language would not be meaningful. That is why semantics focuses on truth conditions, as a way of understanding linguistic meaning. Truth is only one constraint placed on speakers by linguistic meaning, a constraint on the legitimate use of declarative statements. It is just as much a violation of these conventions to say 'Close the door' when you don't want the door closed, or to ask 'Which leg is broken?' when you don't want the information requested, as it is to say something false." Curiously, S. Miller (1986: 138–9; 1992: 42) concludes that conventions need not have alternatives from the premise that there is a convention to tell the truth!

32 Lewis (1975) later strengthened his definition, saying that P uses L provided there is a convention of truthfulness *and trust* in L. I think it is strained at best to treat "trusting in L" as a convention. I can freely choose to be truthful in English, French, or a code of my own. Once you have spoken, however, I am not similarly free to believe that you have said something in English, French, or a code of my own. I am usually compelled by the evidence of your words to believe that you used a particular language. Whether or not I then believe what you have said depends on how knowledgeable and honest I take you to be, and is no more arbitrary than the prior decision. The fact that others are generally trusting in English will be a negligible consideration in making either decision. Cf. Fodor 1975: 72, 78; Bennett 1976: 178–9; and §8.6 of this volume.

fact that English is a living language does not entail that there are honest people. Since English can be used to lie as well as to tell the truth, it seems quite possible for there to be a group of inveterate liars who use English. They use "defendant" to mean "defendant," "was" to mean "was," and so on. In general, they use sentences of English to express propositions expressed by those sentences in English. The only difference is that they generally do not believe the propositions that they have expressed. Second, there seems to be no reason to believe that a community of storytellers could not evolve a distinct language that they use exclusively for literary purposes. One of the customs of this group is to use that language only when telling stories, writing novels, and the like; all other uses are taboo.[33] John Hawthorne described another possibility: that the community of physicists should adopt "instrumentalism," using the language of theoretical physics while believing only the observational consequences of theoretical sentences.

Lewis (1969: 195; 1975: 27) found such cases "bizarre" and "unclear." I see no reason to classify them as anything worse than "merely hypothetical."[34] According to Lewis, "A man lying in Welsh is using Welsh, but he is violating its convention" (1969: 51). He is violating *one* convention associated with Welsh, to be sure. But there are other conventions to which he is conforming – he is using words, phrases, sentences, and constructions to mean what they mean in Welsh.[35] These are the conventions that define what it is to use Welsh correctly. A man who lies in Welsh may be acting improperly, but he need not be speaking Welsh improperly. So the convention being violated is not *essential* to the existence or nature of Welsh. That which is common between the truthful and the untruthful is essential to language. Being truthful is not.

Metaphor and "loose talk" make the same point. As Sperber and Wilson (1986b: 541) emphasize, long stretches of literal speech are the exception rather than the rule. When speaking metaphorically, the speaker is not using sentences to express belief in the propositions that they express

33 Blackburn's (1984: 133) final definition of sentence meaning has no trouble with languages of liars, but does rule that the sentences of this purely literary language mean nothing: "A sentence S means that p in the language of group G if it is a regularity, or a consequence of a system of regularities, with the status of a convention that one who utters S with basic assertive force may be regarded as having displayed that p." To display that p is to produce a natural sign that p. Speakers of a purely literary language cannot utter its sentences with assertive force. Dummett (1973: 2–3) expresses a view similar to Blackburn's.

34 According to Hungerland, lying appears to be taken as normal, and perhaps even right, in the culture of the Barundi, who speak Kirundi.

35 Cf. Laurence 1996: 278–9, 294.

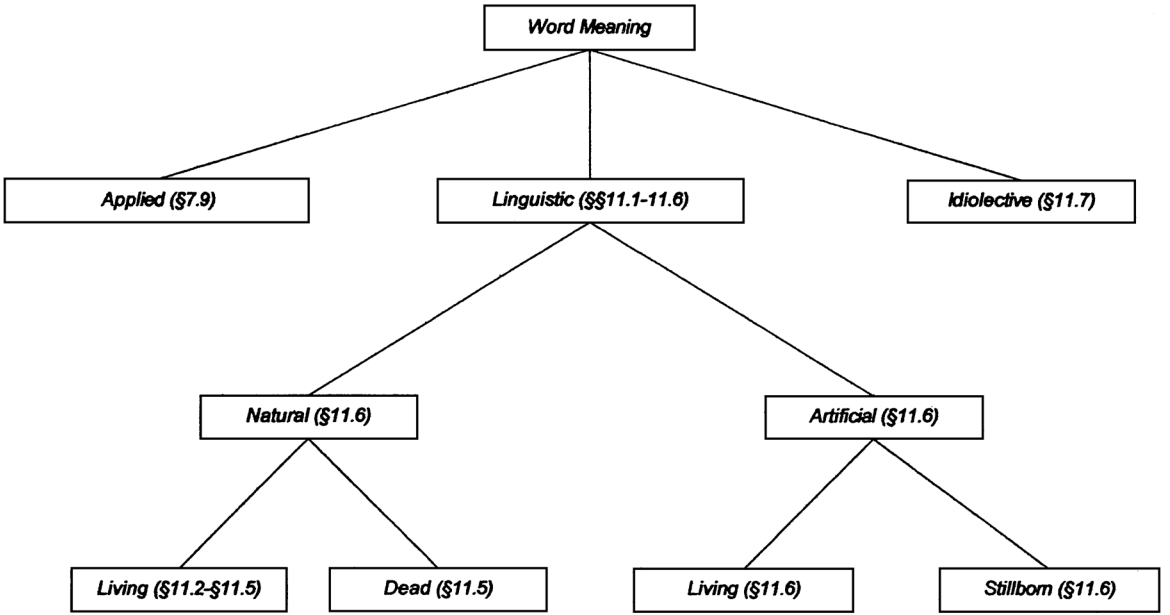


Figure 11.1. The kinds of word meaning

in the language he is using. But he is using the words and constructions in the sentence to mean what they mean in that language. The same goes for the use of language in novels.

Lewis (1969: 177; 1975: 22–3) rightly rejected the suggestion that L is used by P provided that there prevails in P a convention to bestow upon each sentence of L the meaning that L assigns to it. He did so on the grounds that there is no such action as "bestowing a meaning." Actually, there is. When setting up a code, I might say to you, "What meaning should we give to 'run'?" You might respond, "Let's give it the meaning ____." However, such bestowals are relatively uncommon. More importantly, we cannot in this way make it true that a word means something in a living language. It is certainly false that "run" means "run" among speakers of English because it is conventional for them to bestow the meaning "run" on "run." By contrast, expressing beliefs, and using words to mean things, are exceedingly common actions. And it is quite plausible that "run" means "run" because speakers of English customarily use it to mean "run."[36] The absence from Lewis's conceptual resources of speaker meaning and expression severely constrained what he could accomplish.

§11.9 THE KINDS OF WORD MEANING

We have now distinguished, interrelated, and defined several different kinds of word meaning. The first division is based on whether the word means something on an occasion, for a speaker, or in a language. The last is classified on the basis of the kind of language involved. (See Figure 11.1.) Part III will deepen the psychological foundations of the expression theory by defining and clarifying the notion of thoughts and ideas. Part IV will defend the expression theory by rebutting objections to ideational theories in general.

36 Lewis himself defines truthfulness in L rather weakly, which leads to additional trouble. To be truthful in L, according to Lewis (1969: 178; 1975: 14), is to "try not to utter any sentence of L that is not true in L." A speaker can satisfy Lewis's definition: by uttering a batch of sentences that friends assure him are true in English, even though he himself means nothing by uttering them and has no idea what he is saying because he doesn't know English; by using a code in which "Grass is green" means "The guards are asleep," when he knows the guards are wide awake; or by uttering "Snow is white" ironically. Simple ambiguity is problematic (1969: 192–3). Furthermore, if S is truthful in L by Lewis's definition, then S is truthful in every language, artificial or natural, in which the finite set of sentences of L that S has actually uttered are true. And most seriously, "conventions of truthfulness pertain to whole sentences and leave the interpretations of parts of sentences undetermined" (1969: 198; 1975: 20). See also O'Leary-Hawthorne 1990: 1993 and Lewis 1992.

Part Three

Thoughts and Ideas

12

Thought

We have defined cogitative speaker meaning in terms of the expression of thoughts or ideas, and cognitive speaker meaning in terms of the expression of belief. Communication is the effective expression of beliefs, thoughts, or other mental states, and reference is the verbal expression of a certain range of ideas. Word meaning or expression in living languages has been defined in terms of conventional speaker meaning or expression. We turn our attention now to the fundamental notions of thought and ideation.

Belief, desire, and intention have received considerable attention in the philosophical literature. Occurrent thought as a specific propositional attitude has been generally neglected. The notion of an idea has long been suspect, moreover, and ideational theories of meaning are widely regarded as having been thoroughly discredited. It is necessary, therefore, to clarify the relevant notion of thought and ideation, and to respond to well-known objections. I will survey the principal similarities and differences between thought and belief, and stress the distinctive causal role of thought. We will look at the range of terms that express thought in English. While I can present only part of the case here,[1] I hope to make

1 Other parts of the case have been presented in Davis 1981a; 1982; 1987; 1988b; 1992a; 1992b; 1999; and will be presented in my *Belief, Desire, and Thought*. The importance and independence of thought has recently been urged (independently) by Hawthorne ("A Plea for Entertainment"). See also Ryle 1951; 1953; 1958; Geach 1957a; Ginnane 1960; Aune 1967b; Vendler 1972; Hartnack 1972: 543, 547; Malcolm 1973: 13; Carruthers 1989: Chapter 10; Crimmins 1992: 59, 100–2, 125, 127, 152; Fitch 1993; and all the advocates of the language of thought hypothesis cited in Chapter 20. Historically, Ockham (*Ordinatio*) distinguished *apprehension* from judgment. Reid (1785: Essay 4) stressed the fundamentality of thought using the term *conception* (cf. §15.3). In marked contrast to James Mill (1829: Chapter 11), J. S. Mill stressed the independence of belief and thought (1843: 1.5.1; 1869: 402–18).

it evident that conceptually, thought is as primitive as belief and desire, while ontologically it is more general and more fundamental. The often mentioned "belief-desire psychology" is really the "belief-desire-thought psychology."

§12.1 THE COGITATIVE SENSE OF THOUGHT

The word "thought" is at least doubly ambiguous. Like "belief," "desire," and "intention," it suffers from the *act-object ambiguity*. "S's thought that it will rain" can refer either to what S thinks, or to S's act of thinking it. What S thinks, the "object" of his thought, is in this example simply the proposition that it will rain (see §13.3). We will be equally concerned with the act of thinking and its objects. In addition, "thought" is like "meaning" in displaying the *cognitive-cogitative ambiguity*. "Thought" can mean either *believing* or *occurrent thinking*, or their respective objects. We will use "thought" in its cogitative sense to denote occurrent thinking and its objects.

The act-object ambiguity is very well known, but the cognitive-cogitative ambiguity warrants considerable study. Compare the following:

(1) S thinks that p.
(2) S is thinking the thought that p.

Form (1) means that S believes that p. Form (2) means that the thought that p is occurring to S. Form (2) does not entail form (1). Right now I am thinking the thought that Berlin is the capital of France. Now that I have told you what I am thinking, I am sure that you are thinking the same thought. But I do not think that Berlin is the capital of France, and I am sure that you do not either. Moreover, (1) does not entail (2): S may think (i.e., believe) that 5 is a prime number even though he is not at the moment thinking about numbers. I may assert that my wife thinks that 5 is a prime, for example, without knowing what she is doing or thinking about at the moment. Indeed, my statement will be true even if she is asleep or unconscious. Only if my wife suffers a radical memory loss will I have to retract the statement that she believes that 5 is a prime.

Brentano (1874: Chapters 7, 9) and Meinong (1910: xii, Chapters 1, 4, 5; see esp. §17 and §20) similarly stressed the fundamentality and independence of *Vorstellung* ("presentation," "representation," "idea," or "thought") and *apprehending* from judgment. Both used *Vorstellen* to cover sensory as well as conceptual representation. Frege (1918: 7) distinguished "the grasp of a thought – thinking" from "the acknowledgement of the truth of a thought – the act of judgment."

Forms (1) and (2) have different contraries and implications. At least in ordinary circumstances, it is not possible for S to think that p at the very same time that S thinks that not-p. But it is not the least extraordinary for S to be thinking the thought that p while also thinking the thought that not-p. This occurs, for example, when S thinks "It is tautological that p or not-p." It is of course possible for S to think that p at one time and think that not-p at a later time. It follows in that case that S has changed his mind. From the fact that S is thinking the thought that p at one time and the thought that not-p at a later time, it does not follow that S has changed his mind about anything, although it is certainly true that something has changed in his mind.

The fact that beliefs need not be occurrent is often expressed by saying that belief is a *"dispositional" state*. The act of thinking expressed in (2), by contrast, is an *occurrence*, an *act* or *event*. In this respect, thinking resembles saying and uttering more than believing does. That is, (2) resembles (3) and (4), and differs from (1), in implying that *something is going on*, and in being an appropriate answer to "*What is S doing?*"

(3) S is saying that p.
(4) S is saying, "p."

Like statements, and unlike beliefs, it makes sense to ask about occurrent thoughts "*Precisely when and where did S think that thought?*" or "*How often did S do it?*"

Note further that (1) is not simply the frequentive of (2). That is, (1) is not related to (2) as "S plays the piano" is related to "S is playing the piano." "Steve thinks that he will be a millionaire" does not follow from "It is frequently true that Steve is thinking the thought that he will be a millionaire." The latter may be true, for example, solely because Steve fantasizes a lot. Whereas "play" has the same "root meaning" in "plays" and "is playing," the word "think" has different root meanings in (1) and (2). Similarly, (1) is not related to (2) as "x is soluble" is related to "x is dissolving." From the fact that we are all capable of thinking the thought that the moon is made of green cheese, and are disposed to think that thought under certain circumstances, it does not follow that we think (i.e., believe) that the moon is made of green cheese.

Both (1) and (2) can occur in *oratio recta* form:

(5) S thinks "p."
(6) S is thinking the thought "p."

As a freestanding declarative sentence, (5) always has the cognitive sense. "John thinks 'Berlin is the capital of France' " means "John believes that Berlin is the capital of France." But in certain subordinate clauses, (5) is ambiguous. Consider "If John thinks 'Tomorrow is Saturday' during the math exam, then he is not concentrating." This conditional is likely to be true in one sense (where the antecedent describes him as having a certain occurrent thought), false in another (where the antecedent describes him as having a certain belief).

In either sense, "think" is like "believe" in that its sentence-complement need not in any sense be a "direct quote." This is as true when the *oratio obliqua* forms (1) and (2) are used as when the *oratio recta* forms (5) and (6) are used. That is, if "p" and "q" are synonymous, then "S is thinking the thought that p," "S is thinking the thought that q," "S is thinking the thought 'p'," and "S is thinking the thought 'q' " all mean the same thing (see §13.6). There is no difference between thinking "Jimmy didn't win" and "Jimmy did not win," even though there is a clear difference between uttering or saying "Jimmy didn't win" as opposed to "Jimmy did not win." Consequently, when we translate (1), (2), (5), and (6) into another language, the complement "p" gets translated along with the rest of the sentence. Both "think" and "believe" are thus more like the indirect-quote sense of "say" than the direct-quote sense. Forms (1) and (2) do not contrast with (5) and (6), respectively, the way (3) contrasts with (4).

We might try to summarize this point by saying that thinking the thought that p resembles believing that p in being a *propositional* rather than a *sentential* attitude (see Chapter 13). But this formulation is not quite general enough. When it has the cognitive sense, "think," like "believe," must be followed by a declarative sentence. But in its cogitative sense, "think" is like "say" in that its complement can be a quoted interrogative or imperative sentence. As a freestanding sentence expressing belief, "Bill thinks 'Is it going to rain?' " is as ungrammatical as "Bill believes 'Is it going to rain?' " But in the cogitative sense, Bob may think "Will Mary sing?" or "Mary, sing!" just as well as he may think "Mary will sing." If John thinks "Will Mary go out with me?" during the math exam, then he is definitely not concentrating. In the "thought bubbles" of comic strips, which are pictorial renditions of (2) and (6), interrogatives and imperatives appear as readily as declaratives. Nondeclarative sentences appear as readily in the thought bubbles as they do in the "speech boxes" of comic strips. Thought bubbles are ambiguous when they contain declarative sentences, since it is not clear whether occurrent belief or mere thought is being expressed. But

with interrogatives or imperatives, only occurrent thought is a possibility. So whereas the objects of belief are restricted to propositions, the objects of occurrent thought include nonpropositional thoughts.[2]

In the cognitive sense, "think" must be followed by a complete sentence, typically preceded by "that." In the cogitative sense, the quoted expression need not be a complete sentence. Thus *Mary was thinking "After Caesar came, he" – and then she died* is grammatical and may be true. In the cogitative sense, "think" most often occurs with "about" or "of," which must be followed by a noun ("Mary"), noun phrase ("brown cows"), gerund ("going home"), or nominalization ("Mary's loving Jack") that may serve as an direct or indirect object. We have been using "Φ" as a place-holder for such terms, which we call "object nominals" (§6.1).

(7) S is thinking of (about) Φ.
(8) S is thinking of Φ as Ψ.

In both formulas, "thinking" is synonymous with "conceiving." Form (7) will be true if any statement of form (6) is true in which the thought "p" is about Φ. Thus John is thinking of Mary if some thought about Mary is occurring to John. He may be thinking the thought that Mary is pretty, that he loves Mary, that Jane is taller than Mary, and so on. Form (7) does not entail that S is thinking any particular thought about Φ. A fortiori, (7) does not say that S has any particular beliefs about Φ. S can think about the planet Vulcan without even believing that there is such a thing, and can think about winning the lottery without believing that he will win it. Form (8), by contrast, does attribute a particular thought to S. John may be thinking of Mary as his wife. He is doing this if he is thinking the thought that Mary is his wife. He may be thinking this thought whether he believes that she is his wife, or is merely daydreaming that she is. For further discussion of (7), see §12.4.

Objects of occurrent thought constitute what is "*on our minds.*" Thus (7) can be reexpressed using the following locution.

(9) Φ is on S's mind.

"Mary is on John's mind" means that John is thinking about Mary. Skiing is on John's mind if he is thinking about skiing. If "S is thinking about Ψing" is true in the sense that S is thinking the thought that he will Ψ (see §12.4), then we can say:

2 See *Nondescriptive Meaning and Reference* for more on nonpropositional thoughts.

(10) It is occurring to S to Ψ.

That is, (10) means that the thought that *he will* Ψ is occurring to S. Thus "It is occurring to John to go skiing" means that John is thinking about skiing, in the sense that he is thinking the thought that he himself will go. When (10) is in the past tense, there is a suggestion that S actually did Ψ, and wanted to. Neither suggestion is entailed, however. Suppose that Jim has just returned from the supermarket. We might ask him, "Did it occur to you to buy kidneys?" He might answer, "It did, but the idea was repulsive, so I did not buy any."

Many other locutions imply occurrent thought without being equivalent to it. Thus "*imagining* Φ," "*fantasizing* about Φ," "*recalling* Φ," and "*reminiscing* about Φ" describe different ways of thinking about Φ. "S is *considering* (or *entertaining*) the proposition that p" has nearly the sense of (2), but suggests a persistence and purposiveness that is not implied by (2). To consider a proposition is to think it for a period of time in order to form a belief as to whether the proposition is true. (To consider an *action*, like buying a house, is to think about it in order to form a desire to do it or not to do it.) *Pondering*, *reflecting,* and *musing* all imply considering something. A thought may occur to us, though, just for a moment and without such a purpose, in which case we say that the thought *comes (*or *is brought) to mind*, and that it *crosses (*or *enters) our mind*. Finally, to *think up* something, such as an exam question, requires thought directed at producing it, and suggests a certain amount of trial-and-error experimentation. *Thinking out, through,* or *over* a plan, and *calculating* or *working out* a solution, involve even more complex thought processes.

The distinction between cognitive and cogitative thought underlies the distinction between cognitive and cogitative speaker meaning (§2.2). "S meant that p by uttering e" implies that S expressed the sort of thought described by (1) and (5), while "S meant 'p' by the expression e" implies that S expressed the sort of thought described by (2) and (6). Novels provide a common case in which readers end up having occurrent thoughts that are not beliefs, just as they provide a clear case in which the author means things cogitatively but not cognitively.

Philosophers, linguists, and psychologists often use the term "thought" in a generic sense, covering all propositional attitudes and cognitive processes, including both belief and occurrent thought. Think of all the books, series, chapters, articles, and courses entitled "Language and Thought." Philosophers in the Cartesian school used "thought" to denote all conscious mental states whatsoever, as did William James. We are

using the term "thought" very narrowly, however, to denote the sort of propositional attitude expressed by forms like (2), (6), and (7), *as opposed to* (1) and (5). We need some term for this phenomenon, which is similar to, yet notably different from, belief and desire. "Thought" fits the bill nicely. In this sense, the adjective in "occurrent thought" is redundant.

§12.2 BELIEF VERSUS OCCURRENT THOUGHT

In a further effort to clarify the fundamental notion of occurrent thought, and to demonstrate its importance, we will now detail the most important similarities and differences between it and belief. We have observed that believing and thinking are similar in being relations that individuals may have to propositions. Thinking differs in having a broader range of objects, and in being an occurrent event. Belief is a dispositional state that need not be occurrent.

Thinking a thought and believing a proposition are alike, furthermore, in being *intrinsic properties*. Intrinsic properties are those whose acquisition or loss by an object implies a real change in that object.[3] If Bob loses the property of being five feet tall and acquires the property of being six feet tall, then Bob has changed. As a consequence of this change in Bob, Bill may lose the property of being taller than Bob and gain the property of being shorter than Bob. But this change is not a change in Bob. Hence being six feet tall is an intrinsic property, while being taller than Bill is a nonintrinsic property. If Bob is thinking the thought that the sky is blue at one moment, and not thinking that thought the next, then Bob has changed. He has also changed if he believes that God exists one day, and doesn't believe it the next. So thinking the thought that p and believing that p are both intrinsic properties.

Occurrent thought is also like belief in being "*private*": neither can be observed by any of the five senses. This is simultaneously one of the glories and banes of human existence. When I am negotiating to sell something to you, for instance, I am most happy that I can formulate "in my head" the minimum price I will accept without your knowing about it. But I regard it as most unfortunate that I cannot in any way observe your thoughts, which would enable me to know the highest price you are willing to pay. Note that when I seek to "read your mind," or offer the proverbial "penny for your thoughts," I am not just trying to determine your beliefs. I want to know what thoughts are occurring to you. To say

3 Cf. Geach 1969: 66, 71–2, 99; Shoemaker 1984: 207–9; Crane 1991: 6–7.

that thoughts and beliefs are private is not to say that they are unknowable by others. They can be ascertained by standard inductive methods. In one respect, it is easier to detect another's thoughts: if someone says "I was not late," we can usually infer that he is thinking the thought that he was late without worrying about whether he is lying. In another respect, beliefs have the epistemic advantage: we can infer that someone still believes that Paris is in France without having to worry about what is occupying his mind right now.

Ryle and others have argued that thinking is at least sometimes an observable piece of external behavior, as when we "think out loud."

> [W]e do not reserve the title 'thinking' for inner processes. The child, told to think again, is not disobeying if he mutters audibly, 'Seven times seven is forty-nine, nine and carry four.' . . . The architect is thinking out his design for his memorial while studiously arranging and rearranging toy bricks on the carpet; and the composer is not taking a holiday from the labours of composition while his fingers move over the keys, so long as they move in a burdened, searching, tentative and critical manner. (Ryle 1951: 259)[4]

But the child's muttering is not his thinking. While we can observe the child thinking out loud, we cannot even then observe the child's thinking. All we can hear is his muttering, all we can see are the movements of his articulatory apparatus. The child is introspectively aware of his thinking, but not of his thinking out loud. The child must use both introspection and sense perception to tell that he is thinking out loud. The child's thinking out loud may have disturbed the teacher; his thinking could not disturb the teacher. It will generally be true that if the thinking had not occurred, the child would not have got the right answer; it will almost always be false that if the thinking out loud had not occurred, the child would not have got it. For the child could have thought the same thought without doing it out loud. We think out loud by moving our vocal chords, but we do not think by moving our vocal chords. *Pace* Ryle, there *are* two things going on when we think out loud. A child muttering is not thinking out loud unless the muttering is a manifestation of the child's thinking. The muttering must result from and express what the child is thinking, and do so in a particular way. Acoustically identical muttering would not count as

4 See also Wittgenstein 1953: §329–32; Price 1953: 300; Ryle 1958: 394–5; Aune 1967b; Sellars 1969: 104–13; 1979, Chapters 4–5; Vendler 1972: 36–7; Chastain 1975: 235; Devitt & Sterelny 1987: 117; Carruthers 1989: 100–1; Gaulker 1994: e.g., 26. Ginnane (1960: 376) followed Ryle in holding that *all* "intelligent behavior" counts as thought. Contrast Vendler 1977: 60.

thinking out loud if it were produced by direct stimulation of the motor cortex, or were mere recitation of lines from a student play.

McGinn (1997: Chapter 1) observed that sensations differ from propositional attitudes like belief in having "*qualitative content*" in the sense made famous by Nagel (1974): sensations are *conscious experiences*, which are defined by *what it is like* to have them. I know quite well what it is like to have a stomachache, but I can only guess what it is like to have a headache (since I have never had one). There does not seem to be anything it is like to have a belief. I certainly know what it *is* to have a belief. But it seems incorrect to say that I know what it is *like* to have a belief. There is a big difference between believing that Mars has life and not believing it, but the difference is not "phenomenological." This seems clear, given that we retain our beliefs even when we are totally unconscious. McGinn goes on to observe, however, that some mental states have both propositional objects and qualitative content, such as seeing that the sky is blue and being terrified that one will be shot. Thoughts fall into this dual category. There is something it is like to think the thought that the sky is blue. I know both what it *is* to think and what it is *like* to think. We do not have any thoughts when we are totally unconscious. If we are thinking, then ipso facto we are not totally unconscious. It is conceivable, though, that some thoughts are unconscious in the sense that Freud focused on: we may not be aware *that* we have them. Lack of knowledge is at least a logical possibility for any mental state. To say that cats may be conscious or unconscious, and when conscious have conscious experiences, is not to say that they have self-knowledge. *Occurrent* belief is a conscious experience too, because it involves thinking (see §12.5).

Occurrent thought also differs from belief in *not having degrees*. Whereas we can believe in God *more* at one moment than at another, it makes no sense to say that we are thinking the thought that God exists more at one moment than at another. Thoughts do not occur to us in degrees, and we do not think things with differing intensities. We can, though, measure the *frequency*, *duration*, and *latency* of a thought, and the degree of *association* of two thoughts. For instance, we can determine how often Steve thinks about being a millionaire, and how long he thinks about being one on a particular occasion. We can also measure how long it takes him to think of the Russian word for "millionaire" after we ask him, and how often he thinks about being a millionaire when he thinks about his monthly mortgage. It makes sense to ask how long it takes to think something in a way that it makes no sense to ask how long it takes to believe something. While we do not have the technical means to measure precisely how long

a thought takes, we can say that it takes very little time. For we know that thoughts can occur to us "suddenly" or "flash" through our minds. Some are "fleeting." These vague quantitative terms do not apply to belief.

Thought and belief differ extensively in their *causal roles*. One process by which one thought causes another that has no parallel among beliefs is *association*. Two thoughts or ideas are associated when the occurrence of one tends to cause the occurrence of the other in a particular way (Chapter 18). The familiar laws of association, known to Aristotle, link the measurable features of thought just enumerated.

> **The Laws of Association:** *The more recently two thoughts have occurred together, or the more frequently, the more strongly they will be associated.*

Thus salt and pepper are strongly associated among Americans, but not salt and singing. As a consequence, thinking of salt is more likely to prompt an American to think of pepper than of singing, and is likely to prompt the thought of pepper very quickly. The fact that two ideas are associated does not tell us whether S believes or disbelieves any proposition linking them, and strength of association is not related to degree of belief.

Voluntary actions depend on thoughts as well as beliefs, but in different ways. Beliefs combine with desires in a familiar manner to determine what we do. I may go to the supermarket, for example, because I want to buy milk and believe that I can buy milk if I go to the supermarket. The contents of these states form my reasons for acting. The likelihood that I will go to the supermarket, moreover, is directly related to how much I want milk and to how certain I am that I can buy milk if I go. If we use D and B to represent the states of believing and desiring, d and b for degrees of belief and desire, p for probability, M and E for means and ends, and "/" for "given," then these relationships may be summarized informally as follows.

> **The Belief-Desire-Action Rule:** *D(E) + B(E/M) may motivate M; p(M) is directly related to d(M), which is directly related to d(E) and b(E/M).*

Thought, by contrast, is a necessary condition for voluntary action whose content is not a motive. *Forgetfulness*, *absentmindedness,* and *distraction* are major causes of failure to act. No matter how much I want milk, or believe that I can buy milk at the supermarket, I will not go to the supermarket, at

least not on purpose, unless it occurs to me to do so.[5] The likelihood that I will go is also directly related to the likelihood that I will think about going at the appropriate moment. That likelihood, of course, is directly related to how much I want to go to the supermarket ("Objects of desire tend to be on our minds"), which in turn is directly related to how much I want milk and believe that I can get it at the supermarket. It would ordinarily be misleading to say that I went to the supermarket because I thought of going, just as it would be misleading to say that a match ignited because there is oxygen about. But if the other critical necessary conditions of action are evidently satisfied, such a causal statement would be quite in order. All of this is true even though going to the supermarket is in no sense my reason for going to the supermarket. Letting T represent thought, we have:

> **The Thought-Action Rule:** *M cannot occur intentionally without T(M); p(M) is directly related to p[T(M)], which is directly related to d(M).*

These rules are not intended to be precise statements of motivational laws. Nor are they intended to completely describe all of the interesting relationships among belief, desire, thought, and action. I wish only to highlight certain striking differences in the causal roles of belief and thought.

The probability that a subject will think of M is dependent not only on the degree to which that subject desires M, but also on the probability of the subject's thinking thoughts that are associated with M. For example, all year I had the desire to inform reviewers of the outcome when a recent tenure case in our department was decided. The case was decided a month ago, but I forgot about writing the letters until last week, when I noticed a beach house called "Allen's Wood." That reminded me of Allen Wood, which reminded me of his wife, Rega Wood, who was one of the reviewers in the tenure case. I wrote myself a reminder, which I came

5 As a consequence of not introducing thought as a propositional attitude independent of belief, Crimmins gets tied in knots over absentmindedness: "A plausible view of cases of absentmindedness treats them as involving agents who have certain explicit beliefs, but who do not at the time believe their propositional contents" (1992: 59). While this would appear to be a self-contradiction, there is none in saying that I continue to *believe* the proposition even when it does not *occur* to me. It should be noted that while Crimmins's statement is a contradiction in standard English, he is using the term "belief" unconventionally to denote certain concrete particulars (which in my terminology are occurrences or tokens of the thoughts believed).

upon this morning. The explanation of why I spent time this morning writing certain letters would not be complete without reference to my walk along the beach last week and the train of thought that it triggered.

It is commonly noted that our beliefs depend on our reasons for belief. It is seldom observed that what we believe depends in interesting ways on the occurrence of those reasons. The act of discovery depends not just on having the requisite knowledge, but on the often-serendipitous occurrence of that knowledge. For example, for two years after buying a Suzuki Sidekick, I noticed a sort of rattling noise coming from the engine and wondered about the cause. I knew all along that rattle-type engine noise is a symptom of dieseling, and that dieseling can be eliminated by using a higher-octane gas. But it was not until I read a newspaper article on the differences in gasolines while on vacation, that I thought about the noise that my car was making at the same time that I thought about dieseling. As soon as I did, I developed the hypothesis that my Sidekick was dieseling, and that the sound could be eliminated by switching to a higher-octane gas. I did switch, and confirmed the hypothesis. My belief that my car had a rattle-type engine noise and my belief that if a car has such a noise then it may be dieseling did not lead me to believe that my car may be dieseling until the two initial beliefs occurred to me at the same time. As we would say colloquially, I had to "put two and two together."

As a more radical illustration, let me describe a case in which the fact that S has contradictory beliefs does not imply that he is irrational. Suppose that S believes both E and F. Moreover, S believes that E implies H, and that F implies –H. (As a concrete example, let E be "Hitler, Goering, and Eichmann were Germans," let F be "Bach, Beethoven, and Brahms were Germans," and let H be "German culture is sick.") Thus S has contradictory beliefs, and would appear to be irrational. But suppose in addition that E and F simply never occur to S simultaneously. When S is thinking E, it occurs to him that E implies H, but he never recalls that F is true. As a result, he believes H. S continues to believe H, in fact, until F occurs to him. Then it occurs to S that F implies –H, and he believes –H. S does not recall ever having believed H. In this manner, S oscillates unstably between believing H and believing –H. Given the way I have described the case, S can certainly be accused of error and forgetfulness. He just cannot keep all the relevant facts in mind at once, a problem that arises at some point for every thinker, no matter how gifted. But S has at no time been irrational. I would venture to guess that forgetfulness, absentmindedness, and distraction are more common

causes of inconsistent beliefs than irrationality. Note, incidentally, that the case represents a widely overlooked sort of exception to the principle that beliefs are closed under logical implication.

Belief and thought have different "logics." It is irrational, and may be impossible to believe, that today both is and is not Monday. It is neither irrational nor impossible to *think* that thought, however. Indeed, we are doing it now.

The reliability of rules describing the logic of belief can often be improved by adding a suitable thought provision. Take the "addition rule." It is an elementary principle of propositional logic that any proposition P entails any disjunction PvQ containing P. Given that "Paris is in France" is true, so is "Paris is in France or Andorra." It also *tends* to be true that people believe disjunctions containing propositions that they believe. But exceptions arise whenever the subject has never thought the other proposition in the disjunction, Q. If Steve has never even heard of Andorra, we cannot describe him as believing the proposition that Paris is in France or Andorra, even if he believes firmly that Paris is in France. When it is realized that the number of other propositions Q is infinite, and that only a finite subset has ever been thought by S, it becomes apparent that exceptions to an unqualified addition rule for belief are the rule rather than the exception. It is relatively difficult, however, to find an exception to the following principle: if S believes P, then S believes PvQ, *provided PvQ has occurred to S.* If "has occurred" is changed to "is occurring," I can find no exceptions.

The addition rule also illustrates the fact that belief and thought "travel" differently. If S is thinking PvQ, S must also be thinking both P and Q. But a person who believes PvQ will not always believe both P and Q. Similarly, someone who is thinking the thought that humanity will perish if there is a nuclear war must be thinking the thought that there is a nuclear war. But someone who believes that humanity will perish if there is a nuclear war need not believe that there is a nuclear war. We will refer to the general principle here as the second law of occurrence (see Theorem 14.2).

> **The Second Law of Occurrence:** *Thinking a compound proposition implies thinking all of its component propositions.*

No other propositional attitude has this property. Another illustration of the fact that belief and desire travel differently is provided by the familiar fact that "opposites" tend to be associated, which has no cognitive analogue. For example, when positive sex identification is difficult, a person who thinks "That animal is male" is also likely to be thinking "That

animal is female." But under no conditions is it likely for a person to believe both propositions. The logic of thought – the body of principles relating the thinking of compound propositions to the thinking of their components – is as different from the logic of belief as the logic of desire is. Thus the popular view (cf. Fodor 1990a: 14–5) that the causal role of propositional attitudes *in general* mirrors the logical role of their propositional objects, so that PA(P) implies PA(Q) if and only if P entails Q, is at best an overgeneralization based on the preoccupation of philosophers with belief.

Thought also plays an important and distinctive role in emotion. S's emotions may vary widely with no relevant change in his beliefs or desires depending on the thoughts that he is thinking.[6] Thus someone grieving over a lost loved one, or worried about an upcoming exam, may experience at least temporary relief by *getting his mind off* the problem, perhaps by going to an upbeat movie. Vacations work their magic in part because vast networks of associated thoughts normally triggered by our familiar surroundings are dormant for days on end.

Occurrent thought is subject to a much greater degree of *voluntary control* than belief is. John can think about Mary on purpose. He may decide to think about her for the next ten minutes, for example, and may do so by recalling everything he knows about her. When he is tired of thinking about Mary, he might start thinking about Jane. One way in which we direct our thoughts is by asking ourselves questions. Thus if I ask myself "What is the square root of 16?" the answer "4" will almost always occur to me. Of course, people differ in their ability to concentrate, and no one has complete control over his thoughts. Sometimes thoughts about Mary may occur to John when he does not want them to. They may simply "pop into his head." Indeed, he may not be able to stop thinking about Mary, even though he wants very much to think about something else, such as the work he needs to do. He may be so obsessed with Mary that he seeks psychiatric help. If my mind is dull because I am tired, I may ask myself what the square root of 16 is and draw a blank. Our ability to control our thoughts is also limited by our memory. I may want very badly to recall someone's name. But if I have forgotten it, or if it has merely slipped my mind for the moment, I will not be able to. I am still capable of thinking the thought, of course, but I cannot do so at will. Despite

6 Davis 1981a: esp. §6. For further discussion of the role of thought in emotion, see Davis 1982; 1987; 1988b: 462, 473–6; and Stocker 1987. For historical antecedents, see Reid 1785: 460.

such limitations, we do have a considerable amount of control over our thoughts, which enables us to carry out large-scale projects like writing books.

By contrast, we can do little to control our beliefs. John cannot believe that Mary loves him on purpose. It would be most odd if John chose to believe that Mary loves him. But even if he did, he could not carry out his decision at will. And if he is tired of believing that Mary is a woman, he cannot simply change his mind and start believing that she is a frog. *Deciding that p* is a common cognitive process; *deciding to believe that p* is rare. Of course, there is the phenomenon of wishful thinking. Thus John might believe that Mary loves him because he wants her love very badly. But still this would not be a case where John believed something on purpose. Wishful thinking results from desire, but not in the way that voluntary action does. About the only thing we can do to make ourselves believe something is to search for supporting evidence, or to submit to brainwashing. If we already believe something, we can keep ourselves believing it by refusing to consider new evidence or by refusing to reevaluate evidence we already possess.

There is another, more subtle limitation on our ability to control our thoughts. While we are often able to *continue* thinking about something on purpose, we cannot *begin* thinking about it on purpose.[7] For doing something on purpose entails thinking about doing it before doing it. So in order for John to think about Mary on purpose at t, he had to think about thinking about Mary at some time t′ before t. But if at t′ John was thinking about thinking about Mary, then at t′ he was thinking about Mary. Hence he did not begin thinking about Mary at t. Once John is thinking about Mary, however, he can continue to think about her on purpose. Note also that John will carry out his intention to think about Mary by thinking specific thoughts about Mary, such as that she is pretty. But usually John will not think these specific thoughts on purpose. They will simply pop into his mind as a result of his desire to think about Mary. Similarly, when I ask myself what the square root of 16 is in order to think of that number, I will carry out my intention by thinking of 4. But the specific thought that the square root of 16 is 4 will simply pop into my head. I had no prior intention to think that specific thought, otherwise the number 4 would already have been on my mind. In the case of belief, we can neither begin to believe things, nor continue to, on purpose. And

7 This point was observed by Reid 1785: 438; J. Mill 1829: 130; James 1890: Chapter 14, p. 594; and Vendler 1976: 44–5.

we are no more able to form some beliefs about a subject on purpose than to form specific beliefs about it.

Thoughts differ markedly from beliefs in their "*rationality,*" which has both causal and normative components. We typically believe what we do because we have reasons for our belief. Our reasons consist in the evidence that we have for the proposition we believe, and the cause or explanation of our belief is our belief that the evidence obtains. The rationality of the belief depends on how good our reasons are from an epistemological standpoint. It is both unusual and irrational for a man to believe that he won the lottery without having any evidence that he did, that is, without having any reason to believe that he won, such as the fact that the numbers he picked were posted in the newspaper. The fact that if we believe certain propositions, good consequences will ensue, or that we have to believe something in order to reach certain goals, or that it is fun to believe it, is seldom, if ever, a good reason for believing something.

Thoughts are very different. We do not have or need reasons for thinking thoughts in the way that we have and need reasons for believing things. Often, thoughts just pop into our minds. There is nothing irrational about this, although it would be highly irrational if our beliefs arose in the same way. Equally often, thoughts are prompted by events that do not constitute reasons. Thus hearing someone say "Napoleon was short" may cause me to think the thought that Napoleon's invasion of Russia was a disaster. You might wonder how I got from one thought to the other. But you would not assume that the fact that Napoleon was short, or that the speaker said he was, was my reason for thinking the thought that Napoleon invaded Russia. Indeed, you would not assume that I had any reason at all for thinking that thought. There was a reason why I thought what I did, but it was not my reason for thinking the thought. We have reasons for thinking things only to the extent that our thoughts are voluntary, and in that case the reasons we have are reasons for action rather than reasons for belief. Thus if I am thinking about a murder case because I want to solve it, the cause of my thought is my desire to solve the case, and my reason for thinking about the murder case is to solve it. The fact that it is fun to think about something is a perfectly good reason to think about it.

The notion of a *train of thought* is similar to the concept of association. A train of thought is a sequence of thoughts in which one leads to the next causally. Whereas an association is a disposition or tendency of one thought to cause another, a train is an actual causal sequence. A train of thought is typically sustained and directed by a desire to think some

final thought, such as the solution to a problem. Since this thought is the goal of the thinking, it is used to measure the progress of the train. In some cases, however, we simply follow a train of thought wherever it leads us. A train of thought may be *interrupted*, in which case the causal chain is broken by thoughts originating outside the train, or even lost, in which case the subject cannot start it up again after the interruption. The greater the subject's ability to concentrate, the less likely it is that his train of thought will be interrupted or lost. A train of thought may also lead the subject into new territory, to ideas not previously associated with any of the preceding ones.

The notion of a train of thought is also similar to that of an *inference*. An inference occurs when one belief or a set of beliefs causes or sustains another in a particular way. The initial set of beliefs constitutes the subject's reason for adopting or holding the final belief. Some inferences constitute trains of thought, as when I observe that the flag is at half-mast and infer that someone has died. But other inferences do not. It may be true that Kathy infers that 51 is a composite number from the fact that it is divisible by 17 and 3 even when she is not thinking of numbers. And many trains of thought do not constitute inferences, as the "Napoleon was short / Napoleon's invasion of Russia was a disaster" sequence just given illustrates.

Fodor justly criticized associationists for almost completely ignoring the fact that trains of thought often parallel arguments, in which the causes of the beliefs we arrive at are reasons for believing them to be true. But Fodor, and much of contemporary thinking about propositional attitudes, goes too far in the other direction and fails to give associationists their due.

> What is therefore interesting, for our purposes, is that Holmes's story isn't *just* reconstructive psychology. It does a double duty since it also serves to assemble *premises* for a plausible inference to the *conclusion* that the doctor did it with the snake . . . . (Compare the sort of mental history that goes . . . , "Bell-ropes always make me think of snakes, and snakes make me think of snake oil, and snake oil makes me think of doctors; so when I saw the bell-rope it popped into my head that the Doctor and a snake might have done it between them." That's mental causation perhaps; but it's not *thinking*.) (Fodor 1990a: 21)

On the contrary, the associationist bell-rope–snake oil–doctor sequence of events is a very typical process of thought, one, as we saw, that plays an important role in the etiology of action. The process is not *reasoning*, and does not involve the formation of *belief*, so it would not directly interest an epistemologist. But the subject is certainly *thinking*, and one thought is

causing another by a familiar mechanism. The example illustrates another way in which causal relations among one set of propositional attitudes (occurrent thoughts) fail to "mirror" the logical relations among the objects of thought (see also §20.6).

The example also enables us to turn an important epistemological point into a claim for the psychological distinctiveness of thought. It is often observed that there can be no true "logic" of discovery, because one of the key elements of discovery is not a reasoning process. It is standardly observed that while the *testing* of hypotheses proceeds by deduction or induction, the *formation* of hypotheses – our *thinking them up* – does not. Moreover, the way in which we think up a hypothesis has no bearing on the rationality of our holding it, which depends entirely on reasoning that we do after the hypothesis is formed. Thus the theory of the benzene ring was a major scientific discovery, one that has been abundantly confirmed. The fact that Kekulé conceived the hypothesis as a result of having a dream about snakes chasing themselves around in a circle is curious, but does not undermine the hypothesis in any way. The psychologically important point is that unless we think up hypotheses, we will not form beliefs by hypothetical induction. So in another way, thought has an important and distinctive causal role in the formation of belief.

§12.3 THINKING AS THE OCCURRENCE OF THOUGHTS

We will select as primitive the term "thought" in the cogitative, objective sense – the sense in which it is a general term denoting the thought that the sky is blue, the thought "Will Mary sing?" and so on. The thought "p" for any sentence "p," and the thought that p, for any indicative sentence "p," are thoughts. Thoughts in this sense are *events*,[8] specifically, *mental* events. Thoughts are the sorts of events that occur to people or other intelligent beings, and any being to which a thought occurs ipso facto has a mind. For a thought to occur is for some thinking to take place. Thoughts in this sense are event-*types*, however, rather than event-tokens. "The thought that the sky is blue" denotes a type of event, occurring to

8 Cf. Fitch 1993: 470. D. L. Perry's (1967: 91–2) argument that sensations are not events because they are not said to happen, and Cartwright's (1962: 86) parallel argument that words are not events, could be applied to thoughts with equal force. But I think the proper conclusion to draw is that not all events are said to happen. Muscle spasms and births are clearly events, for example, but are not said to happen either. We do not ask "When did the muscle spasm (or birth) happen?" Perry himself argues that thinking is an activity (1967: 197), which surely implies that it is something that takes place.

many people at different times. Given that thoughts are event-types, the act of thought can be defined in terms of their occurrence.

12.1 **Definition:** *S is thinking T iff T is a thought (-type) occurring to S.*

This definition is useful, not because it tells us much about what it is to think, but because it formulates the basic relationship between the act and object senses of occurrent thought, and does so by linking both to the notion of event occurrence.

We will be focusing primarily on the sense of "thought" in which it denotes event-types. But like many other common nouns, "thought" has a closely related sense in which it denotes event-tokens. If I ask "How many primates are there?" I might be asking about how many species of primate there are (about 250), or about how many individual primates there are (many billions). In "How many words are on this page?" the word "word" normally has its token sense; hence we count every occurrence of the word "the," for example. In "How many words are in the English language?" the word "word" normally has its type sense, in which there is only one word "the"; repeatedly typing a word does not increase the lexicon of English. Similarly, suppose that Tom, Dick, and Harry are all thinking "The sky is blue." If I ask "How many thoughts are being thought?" I would most commonly be asking how many thought-types are being thought, and the answer would be "One." But I might be asking how many thought-tokens are being thought; in that case the answer would be "Three." The type sense of "thought" may be taken as primitive. That is, a thought-token can be defined as a token of a thought-type. Any attempt to reverse the conceptual order and define a thought-type as one whose occurrences are thought-tokens is problematic. For any set of thought-tokens could be tokens of more than one type, and some thought-types may never have been tokened.[9]

The verb "thinking" has a parallel ambiguity: John is not only thinking the thought that the sky is blue, he is thinking the token of that thought occurring to him.[10] Definition 12.1 defines the type sense. The token sense of the verb can be defined using the same formula: *S is thinking t (in the token sense) iff t is a thought (token) occurring to S.* Note well, however,

9 Crimmins (1992: 127) reserves "thought" for thought tokens, using "thought map" for types. It is not clear, though, how closely Crimmins's thought–belief distinction matches the one I have drawn.

10 It might be argued that we think tokens in the same sense that we think types, making the verb "think" general rather than ambiguous. As far as I can see, however, nothing hangs on this issue for us.

that when "thought" and "think" have sentential complements, as in (2) and (6), they must have their type senses. In no conventional sense is there more than one thought that the sky is blue, even though it has countless tokens. "Thought that p" must be preceded by the definite article "the," never by the indefinite article "a." Hence "a thought that the sky is blue" cannot be predicated of occurrences of the thought that the sky is blue or anything else.[11]

On the Fregean view, thoughts are abstract objects distinct from mental states, even though some mental states consist in the grasp of thoughts. As even devoted Fregeans have noted, this leaves mysterious what thoughts are, and what it is to grasp them.

> I have conceived concepts as not having any spatiotemporal locations. By themselves, that is, apart from any mental states, they do not participate in causal interactions. Thus the question of the nature of the ontology of concepts is pressing for the approach I am developing in this book. In this chapter I will address the question of how it is possible for mention of concepts, those abstract objects, to play a significant part in the description of the empirical mental states of thinkers . . . . Even among those who most enthusiastically embrace a domain of abstract objects as the contents of propositional attitudes, the answer to the metaphysical question is almost always left in the shadows of obscurity. Frege himself wrote, "[The grasping of a thought] cannot be completely understood from a purely psychological standpoint. For in grasping [the thought] something comes into view whose nature is no longer mental in the proper sense, namely the thought; and this process is perhaps the most mysterious of all" (1979: 145). (Peacocke 1992: 99)[12]

There is no such mystery on my view. Thoughts are indeed abstract, but only because they are event-*types*. Moreover, the thought that the sky is

11 Some philosophers (e.g., Schiffer 1982: 131; Stich 1983: Chapter 11; Crimmins & Perry 1989: 689; Braun 1991: 296; Crimmins 1992: 54–6; Fitch 1993: 477; McKinsey 1994: 304–5) have begun using expressions of the form "a Ψ that p" (where Ψ is a propositional attitude) as general terms denoting tokens of the Ψ that p. Talk of "tokens of the *belief* that p" is especially problematic. It is unclear whether this denotes occurrences of the proposition/thought believed, or cases of believing that p, which are different things. I believe that the questions Schiffer raises – such as "What makes *s* a belief that snow is white?" – are perplexing at least in part because they are ill-formed. Even a statement like "My belief that snow is white is a belief that snow is white" has no conventional interpretation.

12 See also James 1890: Chapter 10, p. 294; Husserl 1900: §1.4.31; Bloomfield 1933: 143; Ryle 1949: 295–6; Gale 1967: 500; Katz 1972: 38–9; Hacking 1975: 49–52; Evans 1982: 104, fn. 24; Cresswell 1985: 56; Allan 1986: 88, 139; Carruthers 1989: 14, 77, 87–8, 90–2; Chierchia & McConnell-Ginet 1990: 58; Schiffer 1990: 260–1; and Nelson 1992: 52–3. Compare and contrast Fodor 1998a: 17–20.

blue is a type of *mental event*. For you to think the thought that the sky is blue is for that mental event-type to occur to you. Hence that type occurs to a subject whenever and wherever the subject is thinking the thought that the sky is blue. That is what it is to "grasp" or "have" a thought. Two people "share" a thought if it occurs to both of them.[13] A thought has no spatio-temporal location itself, although all of its tokens do, just as a sentence type has no spatio-temporal location, although all of its tokens do. Thoughts and other types do have spatio-temporal distributions, though. Thus the thought that the Super Bowl is on television may be restricted to North America, and is pretty much confined to Sundays. Believing, desiring, intending, or hoping that the sky is blue all involve more than the occurrence of the thought that the sky is blue, but we cannot therefore infer that thinking the thought consists in more than the occurrence of the thought, or that the thought is not mental. It is particularly important to note that the fact that "thinking consists in a relation to an objective entity" is no grounds for denying that thinking is an "inner mental process." If thinking is not a mental process, what is? Nor does the objective relation view entail that thinking involves *viewing* thoughts in any sense, or having something in the mind "*aimed at*" the thought (Frege 1918: 26). That would make thinking mysterious.

While it makes no sense to say that some abstract objects have causes and effects (e.g., numbers, geometric lines, sets, units of measure, temporal relations, logical properties), it makes perfect sense to say that event-types (such as homicide, reckless driving, anger, war, hyperventilation, and the fire alarm) do. To say that an event-type (e.g., success) has certain causes (e.g., hard work) and effects (e.g., further success) is to claim that tokens of the type have causes and effects of those types. Similarly, the thought that the flag is at half-mast typically causes the thought that someone has died, because tokens of the first type typically cause tokens of the second. And the idea of salt typically activates the idea of pepper because tokens of the first type activate tokens of the second.

Types, as we have said, are abstract objects. How do types differ from other abstract objects? We can distinguish types from abstract objects like numbers by observing that types are *universals*. Numbers do not have tokens, instances, instantiations, or members. Types differ from universals like properties and sets in that *the properties of a type are typically properties of*

13 Contrast Price (1953: 328–9), who assumed that if concepts were mental they would have to be *tokens*. Compare and contrast Geach 1957a: 13–14 and Crimmins 1992: §3.2.

its tokens.[14] The word "true" has one syllable, as do its tokens. Reckless driving causes accidents because instances of reckless driving cause accidents. Since the thought that John is a man is singular, so are its tokens. The properties of properties, by contrast, are typically not properties of the things that instantiate them (red is a primary color, and apples are red, but apples are not primary colors). And the properties of sets (such as their cardinality and membership) are generally not properties of their members. Types also *occur in other types*. Thus in addition to having countless tokens (individual hydrogen atoms), the hydrogen atom occurs in countless molecules such as water and peroxide. The word "sings" occurs in the sentence "John sings." And the thought that today is Sunday occurs in the thought that if yesterday was Saturday, then today is Sunday. Sets and properties have similar properties. Thus sets can be members or subsets of other sets, and properties can be included in other properties. But while types with other types occurring in them thereby *have a structure*, sets and properties do not. Types of events not only occur in more complex events, but also *occur in places, at times, and to people*. Thus the thought that the Super Bowl is on television may occur to John at 7:00 P.M. in Los Angeles, just as hunger or exhaustion may. Many types can equally well be termed kinds. Thus homicide is a kind or type of murder, anger is a kind or type of emotion, and the lion is a kind or type of animal. Words and thoughts are not conventionally called kinds, but their tokens certainly comprise a distinctive kind of object. I can see no metaphysical reason not to use "type" and "kind" interchangeably, and thus to describe words and thoughts as kinds of things.

Frege maintained that thoughts exist eternally and necessarily.[15] Frege's view may seem absurd on its face. For thoughts cannot exist unless there is someone to think them, and thinking beings are contingent, having existed for only a fraction of the time that the universe has existed. However, the principle that thoughts cannot exist without a thinker is tautological only when we are referring to thought-*tokens*. Frege, however, was concerned with thought-*types*, which are a kind of abstract object. In that sense, there are many thoughts that no one has ever thought, and there may well be thoughts that no one ever thinks. In the same sense, since there is an infinite range of colors, there are colors that no object

14 This characteristic of types and many of the others described in this paragraph are observed by Linda Wetzel in *Of Types and Tokens*.

15 See, e.g., Frege 1884: vi; 1918: 17, 25, 27–8; 1979: 135–8, 148. Contrast Carruthers 1989: Chapter 9; Bealer 1993b: 26–8. Frege allowed that thoughts exist at all times, but preferred to say that they are "timeless."

has ever had, or is ever likely to have. There is an infinite number of sentences that will never get tokened. Frege was particularly concerned with truth, logic, and mathematics. Since the thought that $2 + 2 = 4$ is a mathematical truth, we seem compelled to conclude that it is true at all times and in all possible worlds. The thought was as true a billion years ago as it is today. The assumption that thoughts exist eternally is fully consistent with the fact that we can often date the time when a thought first *arose*. For a thought to arise is for it to be thought for the first time. The Platonist who treats biological species as eternally existing abstract objects can similarly acknowledge that species originate and become extinct. For that concerns when the species is instantiated for the first and last times. The issue of whether types exist eternally and necessarily is a major part of the venerable problem of universals. We will assume that universals exist eternally in all possible worlds, but this assumption is not critical for the development of the theory. The weaker assumption, that thought types may exist even when they are not tokened, will be essential for some developments.

§12.4 THINKING OF OBJECTS

As we noted in §12.1, "think" in its cogitative sense occurs more often with "about" or "of" followed by an object nominal than with "that" followed by a complete sentence. That is, (11) is less common than (12).

(11) S is thinking the thought "p."
(12) S is thinking of (about) Φ.

Form (12) will be true if any statement of form (11) is true in which the thought "p" is about Φ. Thus John is thinking of Mary if some thought about Mary is occurring to John. He may be thinking the thought that Mary is pretty, that he loves Mary, that Jane is taller than Mary, and so on. Form (12) does not entail that S is thinking any particular thought about Φ.

What is it for a thought to be about Mary or about love? We will establish in Chapter 14 that thoughts are complex. They have components that occur together in certain relationships when the thoughts occur. Some of these components are not themselves thoughts. We will use the terms *idea* and *concept* to mean *thoughts or cognitive parts of thoughts* (see Definition 15.1). Thus the thought that Mary should be president is itself an idea, which contains the idea of Mary and the idea of being president. The idea of Mary is the thought-part expressed by the name "Mary"

(Definitions 7.5 and 7.9). Given this framework, it is evident that *a thought is about Φ iff it contains the idea of* Φ.[16] Now a complex event occurs if, and only if, all of its components occur, and occur in the right relationships. It follows that if John thinks of Mary, then the idea of Mary has to occur to John. The fact that the idea of Mary occurs to John does not logically entail that any thought containing that idea occurs to John, although it does appear to be true that ideas occur only as parts of thoughts. Whether this is true or not, it would be self-contradictory to say that the idea of Mary is occurring to John but that John is not thinking of Mary. So we can define thinking of objects in terms of the occurrence of ideas.

12.2 **Definition:** *S is thinking of (about) Φ iff the idea of Φ is occurring to S.*

When "Φ" is a gerund, (12) is generally ambiguous, paralleling the ambiguity of "the idea of Φ" (§7.6). For example, "the idea of skiing" may refer either to the complete thought that one will ski, or to the proper thought-part expressed by "ski." As a result, "John is thinking about skiing" may mean that John is thinking the specific thought that he will ski, or it may have a more general meaning on which it would suffice if John were thinking the thought that skiing is difficult.

We observed in §7.6 that when "Φ" is the nominalization of a sentence "p," "the idea of NOM(p)" is ambiguous. It can refer to the idea "NOM(p)" (a subject concept) or to the closely related, but not identical, idea "p" (a thought). Hence "John is thinking of Mary singing" may be true because John is thinking "Mary is singing," or because he is thinking "Mary's singing brought down the house." For ideas that are not themselves thoughts, such as the idea of Mary or the idea of being president, it appears to be true that *thinking of Φ implies thinking some thought about Φ*. But we cannot insist that this is universally true without generating an infinite regress. If thinking of NOM(p) implied thinking some thought "q" about NOM(p), where the thought "q" contains the thought "p," then thinking of NOM(p) would imply thinking of NOM(q); that would imply thinking of NOM(s), where the thought "s" contained the thought "q," and so on ad infinitum. If "John is thinking of Mary singing" is true because John is thinking the thought that Mary's singing is beautiful, then John is thinking some thought about Mary's singing – that is, some thought containing the idea of Mary's singing. But if "John is thinking of Mary singing" is true because John is simply thinking the thought that Mary is

16 Compare and contrast Harman 1977b: 174.

singing, there is no a priori reason to believe that John must be thinking some further thought about that thought.

In standard English, context (12) is subject to the opaque-transparent ambiguity discussed in §6.2. That is, the nominal following "of" or "about" may be interpreted as either a transparent or an opaque description of the object of S's thought. On the opaque interpretation, but not the transparent, (12) entails:

(13) S is thinking about (of) Φ *as* Φ.

Interpreted transparently, "Oedipus is thinking about his mother" may be true even though he is thinking of his mother as his wife and definitely not as his mother. The opaque interpretation of (12) is true only if a particular idea occurs to S. The transparent interpretation may be true in virtue of a class of coextensive ideas occurring to S. Definition 12.2 holds as long as "of Φ" is given the same interpretation in the definiendum as in the definiens. As announced at the end of §6.2, I will be interpreting psychological descriptions like (12) as opaque descriptions whenever possible.

Context (12) differs markedly from (14), which lacks the preposition "of" or "about."

(14) S is thinking Φ.

In (14), "Φ" can be replaced by a restricted range of nouns and noun phrases, including singular terms like "the thought that John loves Mary," quoted sentences like "'John loves Mary'," indefinite descriptions like "a thought about Mary," and the variable T with thoughts in its domain. But "Φ" may not be replaced by a gerund, a nominalized sentence, or a proper name of anything other than a proposition or thought. Expression (14) is a genuine relational predicate, which entails the existence of Φ. Indeed, (14) is a fully extensional context. Context (12), in contrast, is an intentional context (§6.3), whose truth does not entail the existence of Φ (even on its transparent interpretation). Thinking *about Pegasus* does not entail that Pegasus exists. All that it entails is the existence of the *idea* of Pegasus. Thinking *the thought that Pegasus is a flying horse*, by contrast, entails the existence of that thought. Thoughts must exist for people to think them.

An interesting instance of (14) results when "Φ" is replaced by the proper name of a proposition, such as "Newton's Third Law." Newton's Third Law is the proposition that F = ma. For S to *think Newton's Third Law* is for S to think the proposition that F = ma. For S to think *of* Newton's Third Law is for the idea *of* Newton's Third Law to occur to S.

The idea of Newton's Third Law is a subject concept, not a proposition. The idea that is Newton's Third Law is the proposition that F = ma. It is quite possible for someone to think of Newton's Third Law without thinking Newton's Third Law, as when he thinks the thought "I do not know what Newton's Third Law is exactly, but I remember reading about it in physics," or "Newton's Third Law is the principle that F = ma[2]." Since (14) is extensional, anyone who is thinking the thought that F = ma is thinking Newton's Third Law, even if he does not recognize the proposition he is thinking as Newton's law.[17]

In both (12) and (14), "Φ" may be said to denote the "*object*" *of S's thought*. Thus if John is thinking the thought that Pegasus had wings, then we can say both that the thought that Pegasus had wings is the object of John's thought (because that is what he is thinking) and that Pegasus is the object of John's thought (because that is what John is thinking of). Since it is essential not to equivocate between these two senses of "object of thought," I will use "*relational object of thought*" to denote *what S is thinking*, that is, *the thought that S is thinking*. I will use "*intentional object of thought*" to denote *what S is thinking of.* Relational objects of thought are genuine objects, the things over which the second variable in the relational predicate "S is thinking T" ranges. If T is a relational object of S's thought, then T must exist; moreover, if T = T′, then T′ must also be a relational object of S's thought. Intentional objects of thought, by contrast, cannot be treated as genuine objects, on pain of Meinongian absurdities. Pegasus can be an object of thought even though Pegasus does not exist. And the cube root of 8 need not be the object of John's thought even though the square root of 4 is. "S is thinking of Φ" and "Φ is the object of S's thought" should not be treated as a expressing a relation between Φ and S or his thought.[18] They should be treated only as descriptions of S and what S is thinking.

17 Oddly, "Φ" is seldom if ever an "ism" in (14). We are said to believe, but not to think, materialism, even though we can think as well as believe that everything is material.

18 Contrast Fitch 1987: 114; Yourgrau 1990: 97; Crimmins 1992: 82–92. If thinking of something is called a relation, it is found to be mysterious: "My ideating, my thinking, my feeling and my willing are always in their own peculiar way 'aimed' at something. I ideate *something*, a something that is not the ideating, perhaps a book; my thinking grasps things that are not themselves thinkings, indeed, that do not belong to the mind at all; it grasps them, without in any way drawing them into itself; there is, and there can be, no suggestion of a spatial relation; and yet my thinking 'seizes' those things. . . . A relation, truly, that would be mysterious, nay inconceivable, if we were not so familiar with it from our inner experience! But it is altogether confined to the psychical" (S. Witasek, *Grundlinien der Psychologie*, quoted in Titchener 1929: 198).

Note that "S is thinking *about the thought that p*" is an instance of (12), not of (14). It thus means not that S is merely thinking the thought, but rather that S is thinking *of* the thought as a subject of predication. Bill would be thinking about the thought that smoking causes cancer if Bill were thinking the second-order thought that the thought is true, probable, difficult to prove true, or widely believed but still controversial; that the thought is a subject-predicate proposition, a causal hypothesis, or something believed by the surgeon general, and so on. While the thought that smoking causes cancer is itself an idea, the idea of the thought that smoking causes cancer is a second-order idea, one that is not itself a thought. Thinking a thought T does not entail thinking a second-order thought about T or containing T.

Note also that whereas "thinks" does not have a frequentive meaning when followed by "that," it does when followed by "about" or "of," or when followed immediately by a noun phrase. Thus "John thinks about Mary" means that John frequently thinks about her, that is, it is frequently true that John is thinking about her. Similarly, "John thinks the thought that he will succeed" means that John frequently thinks that thought.

§12.5 OCCURRENT BELIEF

Whereas thinking is an event, and is sometimes a voluntary action, believing is a *state*. We observed earlier that belief is often said to be a "*dispositional*" state. Many philosophers have meant by this that belief could be defined as a behavioral disposition, as a disposition to perform some overt action. This behaviorist view is false. People with particular beliefs do typically have various behavioral dispositions. For example, people who believe that it will rain are commonly disposed to carry an umbrella. But such facts cannot be generalized into a definition of belief. One reason is that a single belief is not universally associated with any behavioral disposition. Thus people who expect rain are disposed to carry umbrellas only because they also believe that umbrellas prevent getting wet. However, by calling belief a "dispositional" state, other philosophers have meant only that a belief *need not be occurrent*, which is undeniable. Belief is a state that may persist for long periods with only occasional manifestations. Most of us have believed that $2 + 2 = 4$ since we were children, for example, but none of us have been thinking the thought that $2 + 2 = 4$ continuously since then. We retain that belief from the time we learn arithmetic until memory fails us or we die. Whereas the total set of propositions that we believe changes only gradually over time, the subset occurring to

us typically changes markedly from minute to minute. A telephone call announcing the death of a loved one may make us change from thinking exclusively about the football game on television to thinking exclusively about the loved one. But even the most momentous phone call will change only a small fraction of our total set of beliefs. Belief could be a neural disposition. Indeed, in §18.4 we will present a connectionist model on which believing that p is a disposition to respond neurally to an occurrence of the thought that p. While the model fits what we know about belief, the characteristic neural response that it postulates has not been identified.

The term *occurrent belief* is often introduced to mean a belief whose relational object is occurring to the subject. Hence:

12.3 **Definition:** *S occurrently believes P iff S believes and is thinking P.*

While the term "occurrent belief" is technical, the concept it denotes is expressed in English by the following common locutions:

(15) It is occurring to S that p.
(16) S is thinking that p.
(17) S is thinking "p."

If it is occurring to Kathy that 5 is a prime number, then she must believe that it is a prime number, and the thought that 5 is a prime must be occurring to her. Thus (10) contrasts with (15): "It occurred to Jim to buy kidneys" implies that Jim thought about buying them, but not that he expected to. In (15) and (16), the "that" restricts "p" to declarative sentences. In (17), "p" may also be imperative or interrogative, but occurrent belief is expressed only when "p" is declarative. Thus S may well be thinking "When will it stop raining?" but this obviously entails no belief. But "S is thinking 'It stopped raining' " implies that S believes that it has stopped. Note that inserting "the thought" in (16) and (17), as in (2) and (6), cancels the cognitive implication. "The novelist is thinking that New York was bombed" entails that the novelist actually believes that New York was bombed. But "The novelist is thinking the thought that New York was bombed" has no such implication: he might simply be thinking up a plot. Note also that while the difference between "S thinks that p" and "S is thinking that p" is the difference between belief and occurrent belief, "S is believing that p" does not stand to "S believes that p" in the same relationship. For "believe" cannot occur in the present progressive; that is, "S is believing that p" is ungrammatical.

The distinction between belief and occurrent belief is analogous to the distinction between standing memory and recall memory. I may remember that President Truman's first name was Harry for the rest of my life, but I will only recall his name on selected occasions. In fact, I may remember that his name was Harry even though the name "slips my mind" on occasion and I am temporarily unable to recall it. There are two ways, consequently, in which I can forget something – say, that I have an appointment. I may lose the belief entirely (it leaves long-term memory), or the belief may not be occurrent at the appropriate time (it leaves short-term memory without returning). One linguistic difference between "believe" and "remember" lies in the fact that while "remember" can be used to express both standing and recall memory, "believe" is never used to express occurrent belief. That is, "Bill remembered that Truman's name was Harry" is ambiguous between standing and recall memory, and may mean that Bill recalled that Truman's name was Harry. "Bill believed that Truman's name was Harry" is not similarly ambiguous, and never expresses occurrent belief. Hence while "S is remembering" is meaningful, "S is believing" is not. It should also be noted that while standing memory implies standing belief, and recall memory implies occurrent belief, the converse implications do not hold. A man witnessing a hit-and-run accident may occurrently believe that the car is green. But he is observing that the car is green rather than recalling that it is, and the belief may not be stored in memory at all. That is, the witness may immediately forget the color of the car.

One kind of decision also implies occurrent belief. "S *decided that* p" means that S formed the belief that p after considering the proposition that p for some time. To decide that p is to *make up one's mind* that p. Deciding that p is a process that results in occurrently believing that p for some (possibly short) time. If the decision is firm, the belief is retained after it is no longer occurrent. Another kind of decision involves the parallel notion of occurrent desire and intention. "S decided *to* Ψ" means that S formed the desire and intention to Ψ after considering whether to Ψ. Hence "S decided to Ψ" implies that it has occurred to S to Ψ, but differs in implying that S intends and therefore desires to Ψ. If the decision is firm, the desire and intention will be retained after it is no longer occurrent.

Definition 12.3 defines occurrent belief in terms of "dispositional" belief. Some reverse the order of priority, defining a dispositional belief as a disposition to occurrently believe: S believes P iff S would occurrently

believe P if P occurred to S.[19] While this formula generally holds, it sometimes fails in both directions. It is quite possible for S to believe something that he would change his mind about if he were to think about the matter again. On the other hand, S may have a disposition to occurrently believe P even though S does not actually believe P. Aristotle, for example, did not believe that neutrons either do or do not have a charge, since he did not have the requisite concepts. But if this instance of the law of excluded middle somehow *had* occurred to him, then he undoubtedly *would* have believed it. Similarly, it may be that if Kathy thought about whether 573 is a prime number, she would swiftly do the necessary calculations and form the belief that it is not. Even though she *would* occurrently believe that 573 isn't a prime if she were to think about it, it does not follow that she actually does believe that 573 is not a prime. We must be careful to distinguish, in other words, between *actual* and *potential* beliefs.[20]

We must also separate the actual-potential belief distinction from that between occurrent and nonoccurrent beliefs. For nonoccurrent beliefs are just as actual as occurrent beliefs. The only difference between an occurrent belief and a nonoccurrent belief is that the content of the former is on the subject's mind at the moment and the latter is not. That is, in the one case believing is accompanied by thinking, and in the other it is not. The notion that "dispositional" belief can be defined in terms of occurrent belief, in the way that solubility can be defined in terms of dissolving, has unfortunately led some to doubt the reality or causal efficacy of belief. Such doubts about dispositions may be misguided, but they are doubly misguided if applied to belief, which is not simply a matter of what one would do under certain circumstances.

Another way of defining dispositional belief in terms of occurrent belief was proposed by Fodor (1987: 22): "I suppose that one's dispositional beliefs could reasonably be identified with the closure of one's occurrent beliefs under principles of inference that one explicitly accepts." This definition fares much worse than the previous definition. First, it appears to give belief even less "reality" than a disposition like solubility. Second,

19 See, for example, Price 1954: 15; 1969: 21, 245; Sellars 1969: 102; De Sousa 1971: 53; Harman 1978: 61–3; Lycan 1981: 141, 153; and perhaps also Goldman 1970: 87. Alston (1967b: 403) advances a similar definition of desire.

20 See Audi 1982 for an extended discussion of this distinction. Crimmins's (1992: §2.3) distinction between "explicit" and "tacit" beliefs is very similar. In my view, what Crimmins means by a "tacit" belief is a potential belief so close to actuality that it is just as if the subject had an actual belief.

the number of occurrent beliefs we have at any given moment is extremely small, whereas the number of beliefs we have is truly enormous. No principles of inference that I accept will lead from the propositions that are on my mind now (when I am thinking about belief, Fodor, my computer, and just a few other things) to all of the propositions that I believe (which include thousands and thousands – about outer space, music, slime molds, sex, tennis, Los Angeles, dirt, math, string, etc., etc.). If Fodor intended the beliefs we had over some period of time to constitute the basis to be closed, then any changes of mind during that period would render the basis inconsistent and the whole approach unworkable. Fodor's proposal seems more appropriate as an attempt to define the *transparent* interpretation of "S occurrently believes that p" on the basis of the *opaque* interpretation (see §6.2). The transparent/opaque distinction cuts across the dispositional/occurrent distinction.

John Hawthorne[21] raised an interesting objection to Definition 12.3, construed not as a stipulation but as an analysis of the concept conventionally expressed by (15). Suppose that P is occurring to a subject who only "tacitly" or "unconsciously" believes P. Let P be the proposition that Sam's neighbor molested him as a child, and suppose that Sam has "repressed" all memory of the incident. Given the assumption that Sam has repressed the memory, when he thinks the thought that his neighbor molested him, he does not think that he believes it. If you ask Sam whether he believes it, he will say no. Indeed, Sam will strenuously deny that his neighbor molested him. Furthermore, Sam may show no inclination to take action against his neighbor, even though he believes that child molesters should be reported and prosecuted and has done so against other individuals he has suspected. He also shows no inclination to warn young children to stay away from his neighbor. Finally, he shows no evidence of being angry at his neighbor, or uncomfortable in his presence. Now, given this background it surely seems incorrect to say that it is occurring to Sam that his neighbor molested him, even though we have stipulated that he unconsciously believes this and is thinking the thought. If this impression is correct, then Definition 12.3 should be modified by replacing "believes" with "consciously believes" in the definiens.

I am not sure that the impression is correct, however. For it is as unintuitive to describe Sam as *believing* that his neighbor molested him as it is to say that it is occurring to him that his neighbor molested him. The reason both claims are unintuitive is that so much evidence has been presented

21 In correspondence. See also Vendler (1977: 55), who focuses on animal belief.

supporting the conclusion that Sam does not believe that his neighbor molested him: Sam denies the proposition, shows no inclination to act in the way that someone who believed the proposition would tend to act, feels none of the appropriate emotions, and does not himself believe that he believes it. This line of reasoning may show that the sorts of cases commonly described using the term "unconscious belief" really are not cases of beliefs (occurrent or nonoccurrent) in the conventional sense of "belief." I do not wish to press such a conclusion, however. What I would insist is that *if* we accept that Sam really does believe that his neighbor molested him, despite all the evidence supporting the claim that he has no such belief, *then* we should also accept the claim that it is occurring to him that his neighbor molested him, despite all the evidence supporting the claim that it is not. That is, if we accept unconscious beliefs, we should accept the idea that it can unconsciously occur to us that something is the case.

§12.6 THE FIRST LAW OF OCCURRENCE

I have been emphasizing that belief and occurrent thought are logically independent propositional attitudes. It is possible to believe a proposition without thinking it at the moment, and possible to think a proposition without believing it. There are countless examples of both possibilities. Nevertheless, belief and thought are not completely independent. First, there may be certain propositions that, for reasons other than logic, cannot be believed without being concurrently thought. Rorty (1970: 420) suggested that propositions about what is happening *here and now* fall into this category. Since these pronouns are *indexical*, they change reference when the time and place changes. So it is hard to see how we could attribute to S the belief that it is raining *now* unless S is now thinking about rain.[22] S may have had the belief for days that it would rain at precisely 10:00 A.M., which happens to be the current time. But that is not the same as believing that it is raining now. There are conversely a large range of propositions that cannot be thought without being believed. I have in mind *self-evident* propositions like "2 = 2."

Second, there is a systematic connection between belief and thought. S cannot believe P unless S has the *ability* to think P. Since Aristotle did not have the concept of antimatter, the thought that antimatter and matter annihilate each other could not occur to him. A fortiori, this proposition is not something that he believed. Furthermore, I believe that S must *actually*

22 See *Nondescriptive Meaning and Reference*.

have thought P at some time or another. That is, S believes a proposition at t only if that proposition occurs to S at or before t. Without prior or concurrent thought, the subject at most has a potential belief.

12.4 **Postulate:** *S believes P only if P is able to occur to S and has actually occurred to S.*

I call this postulate the *first law of occurrence*. We are focusing on belief, but as far as I have seen, the same law holds for all propositional attitudes.

The two clauses of the first law of occurrence are independent. Of course, the proposition that P *is* occurring to S entails that P is *able* to occur to S. But the converse fails. Two minutes ago I was, as usual, able to think the thought that $2 + 2 = 4$, but I did not. Moreover, the fact that P occurred to S at some time in the past does not entail that S is still able to think P. S may have suffered some kind of brain damage between then and now. We normally do not think during sleep, but sleep does not render us unable to think. Indeed, a thought could always wake us up. Hence we retain our beliefs even when we sleep.[23]

Most philosophers have directly or indirectly denied the actual occurrence clause as I have formulated it. Malcolm, for example, argued that a great deal of human consciousness does not involve thoughts.

> I stop my car at an intersection because the light was red. I was aware of the red light and was also aware that the light was red. Did I think to myself, "That light is red"? Probably not. (Malcolm 1973: 20)

I think the evidence that Malcolm was appealing to here was the fact that he probably did not *say to himself* "That light is red." Once we are experienced, driving is much too automatic for that kind of inner speech to occur. But thinking is not inner speech, as we will argue at length in §19.3. Malcolm could not plausibly maintain that he never even thought about the light, or that it never occurred to him that the light was red. Had it never crossed his mind that the light was red, he would not have stopped at the light.

Most philosophers deny the actual occurrence clause by pointing to obvious implications of our explicit beliefs, implications that we never formulate but would immediately assent to if the question ever arose. Schiffer, for example, takes it to be obvious that every philosophy don

23 Comas are a borderline case, I believe. It is unclear whether a comatose person is incapable of thought or merely not thinking, although I am inclined to say the former. It is equally unclear whether a comatose person can still be said to believe things.

at Oxford knows that his maternal grandmother was never married to Benito Mussolini, even though none of the dons has ever thought that particular thought.[24] In favor of the occurrence clause, I offer the following evidence. (1) A statement like "This don knows that his grandmother wasn't married to Mussolini but it has never occurred to him that she wasn't" sounds self-contradictory, and it is hard to think of any situation in which it would be used. The same goes for: "The don thinks that his grandmother wasn't married to Mussolini, but he has never thought that thought"; "He takes it to be true but has never considered it"; and "It has never crossed his mind, but the don is sure that his grandmother was not married to Mussolini." (2) If one is asked, "Does Mary believe that there is molybdenum on Mars," one way to answer the question negatively is to say "She has never thought about the matter." "She has never even heard of molybdenum," which entails that she has never thought about the matter, can be used for the same purpose.

Evidence can also be marshaled for denial of the actual occurrence clause. But I believe close analysis will show that the counterevidence is inconclusive. (3) "This don doesn't know that his grandmother wasn't married to Mussolini" would be a seriously misleading statement, for it would imply that the don is ignorant of whom his grandmother was married to. It would be hasty, however, to conclude that the statement is false, and the corresponding affirmation true. For it would also be misleading to assert "Aristotle did not realize that Neptune perturbs the orbit of Uranus" or "Aristotle did not want Wellington to defeat Napoleon." Yet we could hardly infer that Aristotle did realize or want these things. Furthermore, a nuanced variant of the original statement would not be misleading: "Speaking very strictly, the don doesn't *know* that his grandmother wasn't married to Mussolini." Such a statement would call for an explanation, but it would not wrongly imply that the don didn't know who his grandmother's husband was. (4) If we asked the don whether his grandmother was married to Mussolini, he would answer "No" emphatically and without hesitation. This evidence is inconclusive as well. For asking the don would make the thought occur to him. Hence the truth of the conditional "If asked, he would answer negatively" does not suffice to

24 Schiffer 1972: 36. See also Ockham, *Ordinatio*: 22; Reid 1785: 384–5, 425; Brentano 1874: 201; Meinong 1910: xiii, 9; Titchener 1929: 199–200, 217; Field 1978: 16–17; Harman 1978: 63; Dennett 1978: 104; Lycan 1981: 153; Loar 1981: 248; Devitt 1981: 78; Sperber & Wilson 1986a: 40; and Lance & O'Leary-Hawthorne 1997: §7.6. Contrast Price 1969: 190; Rorty 1970: 420; Lycan 1981: 153, fn. 9 (describing a view attributed to D. M. Armstrong); Sperber & Wilson 1987: 738; and Maloney 1989: 49–52.

prove actual as opposed to potential belief. That is, the conditional may be true because the don *would believe* if the question arose, even though he does not believe because it has not.

(5) Many arguments have been advanced by rationalists, before Chomsky as well as after, for the existence of *innate* knowledge and belief. But the existence or even the possibility of innate belief would constitute a refutation of the principle of prior occurrence only if "innate beliefs" were taken literally to mean beliefs that one is born with. No one perhaps since Leibniz has argued that babies (or fetuses) actually believe things before they have had any experiences or thoughts. What supporters of the innateness hypothesis argue for is the existence of certain structures in the mind or brain that predispose us to forming certain beliefs in response to limited experiences. That is, they are arguing for babies being born with certain innate belief potentials, not with certain actual beliefs (see §17.2).

(6) The most important evidence for denying the actual occurrence clause is that describing the don as believing and knowing that his grandmother was not married to Mussolini does seem intuitively acceptable and natural. However, I believe that such a description is acceptable because we interpret it as a *transparent* description of the don's beliefs, rather than an *opaque* description. The facts cited in (1) and (2), by contrast, hold only when the content ascriptions are interpreted opaquely. In an example used in §6.2 to introduce this distinction, "Oedipus believes that his mother is in bed with him" is true as a transparent description because Oedipus does believe that Jocasta is in bed with him, and she is in fact his mother. Oedipus need not think of Jocasta as his mother, and the thought that it is his mother in bed with him need not have occurred to him. Because this thought did not occur to Oedipus, though, the same statement is false as an opaque description of what Oedipus believes.

The principle that "S believes P" implies "P has occurred to S" holds only when "S believes P" is interpreted as an opaque description. When interpreted transparently, "S believes P" only entails that P is in some sense *implied by* the propositions that S believes in the stricter sense. A weaker law of occurrence therefore holds for transparent descriptions of belief as well as for opaque descriptions: *believing P entails currently or previously thinking some proposition or propositions that imply P.* Since P implies itself, this is a corollary of Postulate 12.4. Our focus has been, and will remain, opaque belief descriptions.

I noted that the first law of occurrence can be generalized to all propositional attitudes. The law holds for disbelief as well as for belief: John disbelieves the proposition that God exists only if John has considered the

proposition at some time, and remains able to do so. The law holds for desire and aversion too: desiring to marry Jane, or being averse to the same, implies that one has at some time thought about marrying her (or about something that implies marrying her), and also that one remains capable of thinking that thought. Ditto for being happy, unhappy, hopeful, fearful, optimistic, pessimistic, and so on. Since all propositional attitudes imply thought, negative as well as positive, thought is the *neutral* propositional attitude.

The claim that thought is the neutral propositional attitude is quite compatible with the further claim that thinking is always accompanied by the possession of, or changes in, other propositional attitudes. It is conceivable, for example, that human beings are psychologically incapable of withholding judgment – that if they think the thought that it is raining in Timbuktu, they must either believe or disbelieve that it is raining in Timbuktu. While left open by the principle that thought is the neutral propositional attitude, this possibility is ruled out by abundantly available introspective and behavioral evidence.

I have been discussing the first law of occurrence in order to clarify the conceptual distinction between belief and occurrent thought. None of the principles relating to meaning or expression in this work depend on accepting this law, however.

13

Sentences, Propositions, and Thoughts

Chapter 12 emphasized the differences between thought and belief, observing that a person can believe something without currently thinking it, and think it without currently believing it. In doing so, however, we presupposed a major point of similarity, namely, that belief and thought can have the same relational objects. Thinking the thought that it will rain and believing that it will rain are different mental states with the same relational object. It is conventional in English and philosophy to refer to the objects of belief as *propositions* (or *statements*). To believe that it will rain is to believe the proposition that it will rain. Not all thoughts are propositions, however. We can think "Will it rain?" but this thought cannot be an object of belief or disbelief, and cannot have a truth value. We will define propositions as "declarative" thoughts, and will briefly examine the contrasting category of nonpropositional thoughts. After setting out our standard method of referring to propositions, we will defend the corollary synonymy criterion of propositional identity against the Mates objection and others.

Thoughts, propositions, and beliefs are all closely related to sentences. The fact that thoughts are expressed by sentences, and that speakers use sentences to express thoughts, played a major role in the theory of meaning presented in Parts I and II. We exploit secondary conventions of ideo-reflexive reference (§7.6) in order to use sentences to refer to the thoughts they express. The sentences that a speaker utters arguably constitute the best evidence we have about the thoughts that he is thinking. Some have gone so far as to treat thoughts as sentences, and thinking as a relation to sentences. On some views, thinking is saying-to-oneself, an internal form of utterance. We can certainly use the similarities between sentences

and thoughts to gain a basic understanding of what thoughts are like. However, there are essential differences, and recognizing the dissimilarities between sentences and thoughts is critical in understanding the nature of thoughts. So this chapter will begin to explore the scope and limits of the sentential analogy. We will also observe that thinking cannot be treated as a relation to the situations that make thoughts true, nor to the sets of possible worlds in which the thoughts are true. *Thoughts, propositions, sentences, situations,* and *world sets* all play essential roles in the system that we are developing. But the roles are different, and need to be carefully distinguished.

§13.1 THOUGHTS VERSUS SENTENCES

There are many important similarities between sentences and thoughts. First, both are *representations*. The thought "The sky is blue" and the sentence "The sky is blue" are both about the sky, and both represent the fact that the sky is blue. Second, sentences and thoughts are both capable of being *true or false*. The thought and the sentence just mentioned are both true because, in fact, the sky is blue. Sentences are true or false in virtue of their *meanings*, thoughts in virtue of their *contents*. Third, sentences and thoughts are both *attitudinally neutral* representations. The subject that produces them may or may not believe that they are true, may or may not desire that they be true, and so on. Fourth, such representations are *abstract* in a way that pictorial and iconic representations are not. A portrait of President Washington that represents him as having false teeth will also represent him as having countless other characteristics: a smile or a frown, open or closed eyes, long or short hair, and so forth. The sentence "President Washington had false teeth" represents only the fact that Washington had false teeth, by contrast, and the same goes for the thought that the sentence expresses.

Fifth, both sentences and thoughts are not only *structured*, but also have a *phrase* or *constituent* structure. As we will show in Chapter 14, just as a sentence is an expression consisting of expressions, so a thought is an idea consisting of ideas. A sentence cannot be uttered or produced unless the words composing it are uttered or produced, just as a thought cannot occur to a subject unless the ideas composing the thought do. Moreover, both expressions and ideas may have any number of levels of constituents. Just as a sentence may contain two other sentences, each of which contains a noun phrase and a verb phrase, which consist of a number of individual

words, so a thought may contain two other thoughts, each of which contains a subject concept and a predicate concept, which consist of a number of individual ideas.

Sixth, both sentences and thoughts have *systematic* or *generative* structure. These terms, used interchangeably, customarily convey two distinguishable features in addition to constituent structure. (a) Different thoughts and sentences may have the same constituents in different relationships, or different constituents in the same relationship. Moreover, an infinity of different thoughts and sentences can be generated from a finite number of unstructured components and structural relationships. (b) What a complex word or idea represents depends on what its components represent and their mode of combination. The thought that John loves Mary represents *John* as loving *Mary* because the thought contains the idea of John and the idea of Mary, rather than, say, the idea of cream cheese and the idea of Jell-O. The thought that Mary loves John represents *Mary* as loving *John*, despite having the same components, because they are combined differently. The generative nature of ideational structure will be developed a bit more formally in *Nondescriptive Meaning and Reference*.

Seventh, the *compositionality* of natural language semantics means that the idea expressed by one linguistic unit – whether it be a word, phrase, clause, or sentence – is generally a constituent of the ideas expressed by the larger linguistic units containing it. Thus the idea expressed by the word "Mary" is a constituent of the idea expressed by the phrase "loves Mary," which is a constituent of the thought expressed by the sentence "John loves Mary." Compositionality means that the structures of sentences and thoughts are similar not only in kind, but also in detail. Indeed, their structures are *isomorphic* to a high degree.

Finally, thinking is like saying or uttering in many respects. Both are actions, whose objects can be represented by using quoted sentences. Thinking "Mary is pretty" is closely related to *saying to oneself* "Mary is pretty." Given these similarities, the convention has developed in comic strips and other pictorial media of putting sentences inside different geometric figures – speech "boxes" versus thought "bubbles" – to distinguish what the subject is saying from what he or she is thinking.

In short, thoughts are like sentences in being attitudinally neutral, abstract, generatively constructed representations with truth values and highly isomorphic constituent structures. Thinking resembles inner speech in consisting of the occurrence of sentencelike mental representations.

The Sentential Analogy: *Thoughts are sentencelike mental representations, and thinking is like inner speech.*

When employed with caution, the sentential analogy provides a useful characterization of thought.

Philosophers, linguists, and psychologists have often claimed not just analogy but identity. Many contemporary cognitive scientists characterize thoughts as "sentences in the language of thought" (Chapter 20). Many introspectionist psychologists have treated thought as inner speech (§19.3). And the logical positivists treated propositions and thoughts as "logical constructions" out of sentences, reducing talk about propositions to talk about sets of synonymous sentences.[1] It is critical in understanding the nature of both thought and language to recognize that *thoughts are not sentences*. While it is a platitude that sentences express thoughts, it makes no sense to say that anything expresses a sentence, or that thoughts express something. Sentences are used to assert, order, amuse, and so on; thoughts are not. Sentences can be written or spoken, and it is reasonable to identify them with sequences of sounds or letters, or types of such. Thoughts are not composed of sounds or letters, and the "production" of thoughts does not count as talking or writing. Thoughts are mental events, and thinking is their occurrence. Sentences are not mental events, and their occurrence does not constitute thinking. One-word sentences are possible. Even though such sentences have no relevant structure, thoughts are never structureless. We speak of sentences as being ambiguous or unambiguous, and as being synonymous with, translations of, or paraphrases of other sentences. These terms do not apply to thoughts.

The relationship between sentences and the thoughts that they express is not even a one-to-one correspondence. Consider the following.

Every man is an animal.
All men are animals.

We have here two different sentences. While both are from the same language, their (surface) grammars are very different. One has a singular subject and predicate, the other has a plural subject and predicate. In fact, these two sentences do not have a word in common. Nevertheless, the two sentences are completely synonymous, expressing exactly the same

1 See, for example, Russell 1940: 208–9; Ayer 1952: 88. A variant is found in Burks: "A sentence may be called a statement [= proposition] when it is regarded as representing any sentence that expresses the same statement as it expresses" (1977: 3).

thought. Similar comments apply to "Bill's accuser is mad" / "The accuser of Bill is mad," "The square root of four is greater than or equal to two" / "$\sqrt{4} \geq 2$," and so on ad infinitum. Consider next:

It is raining.
Il pleut.
Es regnet.

Here we have three different sentences from three different languages – English, French, and German. Since the sentences have the same meaning, they express the same thought. Finally, consider:

Flying planes can be dangerous.

Here we have one sentence. But because it is ambiguous, it can be used to express two different thoughts: the thought that the act of flying planes can be dangerous, and the thought that planes that are flying can be dangerous. In addition to establishing that the expression relation is not a one-to-one correspondence, the facts reviewed in this paragraph also show that *thoughts are grammar- and language-independent*. In saying that thoughts are grammar-independent, I of course mean that they are independent of the "*surface*" or constituent structure of any sentence expressing them. The relationship between thoughts and "deep structure" may be closer, depending on the exact definition of deep structure, which varies from linguist to linguist. On one view, the structure of the thought expressed by a sentence *is* its deep structure.

The distinction between thoughts and sentences provides a simple way to explain the classic Sachs (1967) experiment.[2] Subjects listened to a paragraph, and then later were presented with a sentence and asked to judge whether it was the same as any sentence in the paragraph. Sachs found that subjects often mistakenly thought that they remembered synonymous sentences, but rarely thought that nonsynonymous sentences had occurred. One natural explanation is that the subjects were paying more attention to the thoughts expressed than to the sentences used to express them, and thus remembered the thoughts more clearly than the sentences.

In claiming that thoughts are language-independent, I do not mean to deny that there are many causal and other empirical relationships between

2 Stillings et al. 1995: 27. Precursors were performed by the Würzburg school early in the century. See Humphrey 1951: 61–3.

the thoughts or beliefs that an individual possesses and the language that he or she learns.[3] Evidence cited by Whorf (1956) and others suggests that there are many forms of linguistic relativity. For example, people who have a single term for a given color (describing it as "pink" rather than "pale red") may be better at recognizing and remembering it. And speakers whose expressions for two phenomena are grammatically related may see greater similarity in the phenomena than others do (as in Whorf's examples of how the Shawnee would similarly express the, to us, dissimilar phenomena of pulling a branch aside and having an extra toe, and how we use "have" to express the intrinsically dissimilar phenomena of possessing and giving birth to). Finally, there are the well-known facts that some words are hard to translate, and that thoughts about quarks or imaginary numbers are unlikely to occur in subjects without some language. But none of the Whorfian evidence shows that sentences cannot be ambiguous, or that very different sentences cannot express the same thought. (Nor does it establish the linguistic determinism that Whorf claims in unguarded moments.)

The sentences of any given natural language express particular thoughts in virtue of the conventions constituting that language (see Part II). As languages change, the sentences of the language may come to express different thoughts. Sentences may also come to express different thoughts by stipulation. Thus it is perfectly possible to invent a code in which "The flowers are blooming" expresses the proposition that it is raining. This usage would be unconventional, to be sure, but nonetheless meaningful. Given that it depends on variable factors such as convention and stipulation, which depend on intention, *any relationship between sentences and thoughts is contingent*, and contingent metaphysically as well as logically. In a very real sense, every sentence could express a different thought, and every thought could be expressed by a different sentence. As a consequence, even though thought content and sentence meaning are similar properties, *sentences have their meaning contingently, by convention or stipulation; thoughts have their content essentially, by nature.* The sentence "Grass grows" could have meant "Snow falls," but the thought "Grass grows" could not have had the content "Snow falls."

The power of stipulation is regularly used to enrich languages. When new phenomena are named, we are able to express thoughts that were previously inexpressible in that language. This gives rise to another respect in which thoughts are language-independent: *at any given time, many thoughts*

3 For a brief review, see P. T. Smith 1994.

are inexpressible in any existing language.[4] Thus the thought that solid-state televisions are electronic is true and has always been true. But it was inexpressible in ancient Greek or any other language of the period. Since it is reasonable to assume that human beings will continue to discover new phenomena, we can infer that there are also thoughts that are inexpressible in any language existing today. When these propositions are thought, language will be modified to express them. More controversially, it is plausible that infants and animals have thoughts that will never be precisely expressed in any language, assuming that language remains restricted to older humans.[5]

§13.2 THINKING NOT A RELATION TO SENTENCES

Given the distinction between thoughts and sentences, it follows that *thinking cannot be characterized as a relation to sentences.*[6] From the fact that S thinks the thought that the ball is red, it does follow that S is related in a certain way to the sentence "The ball is red." Our standard *oratio recta* and *oratio obliqua* methods of ascribing thoughts exploit this relationship. Thus we may say "S is thinking 'The ball is red.'" But the implied relationship is neither essential nor fundamental to thinking. For the relationship presupposes the existence of the sentence, whereas thinking the thought does not. And the connection depends on the meaning of the sentence, which could have been different. If "red" had meant "blue," then a person who thinks the thought that the ball is red would be related in the indicated way to "The ball is blue" rather than to "The ball is red." (In that case, of course, we would say "He is thinking the thought that the ball is blue" rather than "He is thinking the thought that the ball is red.") Ambiguous sentences show, moreover, that reference to an interpretation, which determines the thought expressed by the sentence, will

4 Cf. Weiskrantz 1988; 1997; Lance & O'Leary-Hawthorne 1997: §7.6. This is a difficulty for Schiffer's (1987a) attempt to avoid treating propositional attitudes as genuine relations to propositions by using substitutional rather than objectual quantifiers. Contrast also Hartnack (1972: 551), who maintained that thoughts are necessarily expressed in language.

5 Contrast Stich 1983; and see §20.4, this volume.

6 Compare and contrast Carnap 1947: §14; Church 1950; Pap 1955; 1957; Chisholm 1958; Quine 1960: 219; Kneale & Kneale 1962: 604–8; Hunt 1962: 13; Davidson 1968; 1975: 167; Sellars 1969: 104; Hartnack 1972: 546; Danto 1975: 16; Loar 1976b: 147; 1981: Chapter 7; 1987: 180; Wettstein 1976; Field 1978; P. S. Churchland 1980; Fodor 1981: 187–203; Schiffer 1982: 139–45; 1987a: Chapter 5; Stich 1983: 38–40, 73–88; Lycan 1985: 87; Katz 1986a: 68–71; Richard 1989: 326–8; Fodor & Lepore 1992: 139–42; Lance & O'Leary-Hawthorne 1997: §7.6; and J. G. Moore 1999b. See also §19.3 of this volume.

be necessary to determine which relation to a sentence is implied by a thought statement. Under one interpretation, S may stand in the indicated relation to "Mary slept on the bank"; on another, S may not. This makes it clear that it is the thought expressed that is important, not the sentence that happens to express it. Finally, since there might be a proposition that is not expressible in any existing language, it is possible for a person to think something even though he is not related in the indicated way to any sentence. This is a possibility, of course, that I cannot illustrate.

Stich (1983: Chapter 5) proposed a sophisticated, functionalist variation on the Carnapian analysis designed to avoid such problems. Stich focused on belief, but a similar analysis of thought is no less plausible.

"S thinks 'p'" means that S is in a thought state similar to the one which would play the central causal role if A's utterance of "p" had had the typical causal history.[7]

"A" here denotes the person asserting "S thinks 'p'." It is counterintuitive to think that when describing someone else as thinking something, we are also referring indexically to ourselves, but we will let that pass. Stich assumes that "the typical causal history" refers uniquely (without stipulation) to the etiology of a "sincere, straightforward assertion." This assumption too can be challenged, on the grounds that the use of declarative sentences in fiction and indirect discourse is also quite typical. But this is not a problem when thought rather than belief is being analyzed. A similar understanding will have to apply so that Stich's analysis does not assign an idiosyncratic meaning to the English sentence "S thinks 'p'" when A typically uses "p" with a meaning it lacks in English.

The defects of the sort of analysis that Stich proposed are many and serious. First, Stich's analysis does not avoid the problem that there might be a proposition that is not expressed by any sentence "p" in any language. Second, many substitution instances of "S thinks 'p'" will never have been uttered by any speaker A. Any number of these might be true or false. But Stich's analysis cannot be applied to them, because it presupposes utterance. Third, by adopting the Davidsonian (1968) gambit of referring demonstratively to an actual utterance or token of "p," Stich thought that he could avoid the interpretation problem noted earlier. But the problem just reemerges in another form in light of the counterfactual clause. For there is more than one counterfactual case in which my utterance of "p" had a typical causal history whenever "p" is ambiguous. My utterance of

7 See also Churchland & Churchland 1983: 303.

"Jill went to the bank" could have been an utterance of a sentence meaning "Jill went to the river bank," or an utterance of a sentence meaning "Jill went to the commercial bank." Different thought states would typically have been involved in those two cases. And since A could have been using a code, the assumption that she was using "p" to make a sincere and straightforward assertion implies nothing about what thought A expressed. Fourth, similar but slightly different problems arise with indexical sentences, which are conventionally used by different speakers to express different thoughts even though they are not ambiguous. When Stich thinks "I am right," he is not thinking the thought that *I* am right, even though (a) that is the thought that *I* would typically use "I am right" to express, and (b) the thought that Wayne Davis is right is *similar* to the one that is mentioned in (a).

Fifth, several thoughts are centrally involved in a typical, sincere assertion of "p": not only the thought that p, but also the thought that "p" means that p, the thought that one will (or should) assert that p, the thought that it is not offensive to assert that p, the thought that one's audience will come to believe that p if one utters "p," the thought that one is providing useful information, and so on. Stich's (1983: 81) criterion of "centrality" does not discriminate among these thoughts. Sixth, even if we stipulate that the thought that p plays the central causal role in utterances of "p," many thoughts other than the thought that p are similar to the thought that p. Suppose we are concerned with the thought that grass is green, and assume that it plays the central causal role in typical utterances of "Grass is green." Similar thoughts include the thought that grass is blue, the thought that snow is green, and even the thought that grass is not green. Hence Stich's analysis rules that someone who is thinking "Grass is blue" can, in appropriate contexts, be said to be thinking "Grass is green." The defects of Stich's analysis are thus much like those of Smart's (1959) infamous analysis of "orange afterimage."

Thinking the thought that p also differs markedly from thinking the thought that the sentence "p" is true. A Frenchman may think the thought that the sky is blue without thinking the thought that the English sentence "The sky is blue" is true; for he may think a lot about the sky but little about English. Conversely, our Frenchman may think the thought that the sentence "This shirt is maroon" is true without thinking the thought that the shirt is maroon, because he is mistaken about the meaning of the English word "maroon." Alternatively, he might take the sentence to be true without even knowing what it means, because a reliable authority told him that the sentence was true.

Many philosophers have focused on a useful generalization about belief that limits itself to competent speakers of the language used to describe what people are thinking. The parallel principle for thought may also be called the *"disquotation principle."*

(1) Normally, S is thinking the thought that p iff S is thinking the thought that the sentence "p" is true, provided S is a competent speaker of the language to which "p" belongs, and provided "p" is nonindexical and unambiguous.[8]

Since we are using English, S must be a competent speaker of English. When (1) is translated into French, it applies only to competent speakers of French. The disquotation principle for belief is plausible in part because of the dispositional nature of belief. A competent speaker of English who believes that Mars is lifeless is likely at some point to have reflected on the sentence "Mars is lifeless," formed the belief that the sentence is true, and retained that belief. But it is phenomenologically implausible that people thinking about life on Mars are *normally* thinking about sentences. It is particularly implausible that whenever an Englishwoman thinks a thought, she normally thinks about *every* sentence that expresses that thought in English. The disquotation principle for thought is, I believe, false.

Even if it were true, the disquotation principle could not be used to support the claim that thinking is a relation to sentences. First, the principle is restricted to competent speakers of the metalanguage, and to thoughts expressible by unambiguous, nonindexical sentences. Given that the disquotation principle characterizes such a special case, it cannot tell us what it is in general to think a thought. Second, the principle tells us only what is *normally* correlated with a thought in a special case. The disquotation principle would be patently false if that "normally" were omitted. For there are exceptions to the rule. For example, competent speakers of English do not know the meaning of every word in English, and hence need not know that a sentence expressing something that they believe is true. Competence in English, furthermore, is compatible with a limited number of mistakes about what English words mean. Thus a woman who is thinking the thought that Dr. Jones is an entomologist may mistakenly think that "Dr. Jones is an entomologist" is false, because she believes that "entomologist" means "one who studies words." The requirement that "p" be unambiguous is designed to protect the right-left conjunct of (1) from cases in which S interprets "The river rose over the bank" to mean that the river rose over the riverbank, while we

8 See, e.g., Kripke 1979: §2; D. Sosa 1996; Moore 1999a.

(the speakers) use it to mean that the river rose over the savings bank. In that case, S might be thinking that the sentence is true even though he is not thinking the thought that the river rose over the savings bank. But the proviso does not exclude all of the problems. It is hard to be sure that any particular sentence is unambiguous, but let me assume that "2 + 2 = 5" is. Now suppose we ask S to *imagine* that "2 + 2 = 5" has a different meaning that would make it true. Suppose S succeeds in imagining this, and manages to put the conventional meaning of the sentence completely out of his mind. Then S is thinking the thought that "2 + 2 = 5" is true without thinking the thought that 2 + 2 = 5.

Kripke's (1979) puzzle demonstrates the possibility of a deeper class of exceptions. He imagines that Pierre grows up in France, and learns only French. He learns secondhand about the city called "Londres" in French, and forms the opinion that it is pretty. Later, Pierre moves to a city in England that happens to be London, and picks up English from his neighbors. He learns firsthand about the city he is living in, and forms the opinion that it is ugly. He does not realize that he is living in the city he calls "Londres." Now suppose that after learning both languages, he thinks the thought that "Londres est jolie" is true and the thought that "London is ugly" is true. Since Pierre believes that both sentences are true, and does not seem to be guilty of any lapses of rationality, it would be hard to believe that he is thinking contradictory thoughts about London. We must therefore conclude that the disquotation principle fails in at least one of the two cases. Kripke found the case profoundly puzzling, because he could not see any basis for deciding which application of the principle fails and which succeeds. That is in part because he assumed that "as an English speaker," Pierre "does not differ at all" from his neighbors. This conflicts with Kripke's assumption that before he moved to England, Pierre "differed not at all" from his French neighbors. If so, then Pierre already had the concept of London when he began learning what "London" means. He did not realize that "London" expresses the concept of London too, and did not realize that the city around him was London (the city he would call "Londres"). He ended up using "London" to express a different concept, enabling him to believe that "Londres is London" is false. He believes the negation of the identity proposition in which the two individual concepts are those he expresses using "Londres" and "London." Since the identity proposition contains two different individual concepts, its negation is consistent, and Pierre could rationally believe it. Pierre is thus like Peter in the Reagan case of §6.2. So Pierre, if he was just like his French neighbors, must differ

in significant respects from his English-speaking neighbors. Kripke also seems to assume, although not explicitly, that what Pierre is thinking has to be *determined* by the objective, observable evidence we use to decide what he is thinking. But this is simply verificationism. Any inference from observable behavior to inner mental states is inductive and defeasible.

There is a final reason why the disquotation principle cannot be used to support the thesis that thinking is a relation to sentences. The principle correlates thinking a thought with thinking the thought that a sentence is true. It tells us nothing about what it is to think *that* thought.

The identification of thought with inner speech or with "sentences in the language of thought" will be criticized in §19.3 and Chapter 20, respectively.

§13.3 PROPOSITIONS AS THOUGHTS

We have taken as our primitive term "thought" in the sense in which it means an object rather than an act of occurrent thought, and a thought-type rather than a thought-token. The fact that sentences express thoughts was fundamental to the theory of meaning presented in Parts I and II. We also say that declarative sentences express *propositions*. "Proposition" has long been a term of art for philosophers and logicians. On the most prevalent definitions, "proposition" is taken to mean *interpreted declarative sentence, set of possible worlds,* or *situation*. In none of these senses is it true that sentences *express* propositions, in the sense of "express" that we have defined. For sentences, world-sets, and situations are not mental events whose occurrence or nonoccurrence can be indicated by performing an action. We will use the term "proposition" instead in the sense in which it denotes a *relational object of belief or disbelief*. The proposition that the sky is blue is *what we believe* when we believe that the sky is blue. Tautologically, to believe that p is to believe the proposition that p. Propositions in this sense, I submit, are thoughts. What we are thinking when we think the thought that the sky is blue is exactly the same as what we believe when we believe that the sky is blue. *Thinking the thought that p* is the same process as *conceiving the proposition that p*. Expressing the thought that p is the same property as expressing the proposition that p.

While all propositions are thoughts in this sense, not all thoughts are properly called propositions. One can think "Will it rain?" but that thought does not count as a proposition. The thought "Will it rain?" is not the sort of thought that can be believed or disbelieved. Some thoughts cannot be believed or disbelieved because they are too complex, or because

we do not have the requisite concepts, or because the evidence compels us to suspend judgment. The reason the thought "Will it rain?" cannot be believed or disbelieved, by contrast, is that it does not have the right kind of structure. The thought is "interrogative," we shall say, rather than "declarative." Propositions are declarative thoughts.

13.1 **Definition:** *T is a proposition iff T is a declarative thought(-type).*

In describing thoughts as declarative, we are not assuming that thoughts are sentences, nor are we saying that declarative thoughts have the same structure as declarative sentences. We are assuming that declarative sentences express thoughts with a common and distinctive structural feature, and are using the term "declarative" in a second sense to denote that feature. Propositions are thoughts with that feature.

Propositions are often defined as thoughts *capable of being true or false*. Whereas the thought "It will rain" is true or false, the thought "Will it rain?" cannot be so evaluated. This definition is approximately correct, but is problematic for two reasons. First, on the Strawsonian view of presupposition that I accept, propositions with false presuppositions, like "The present king of France is bald," are neither true nor false. Hence propositions with necessarily false presuppositions are not capable of being true or false. An example would be "The prime number between 3 and 5 is bald." Second, propositions like "This sentence is false" are "paradoxical." Since the assumption that they are either true or false leads to a contradiction, we have to conclude that they cannot be either. Both thoughts we have mentioned count as propositions on Definition 13.1, because they have a declarative structure. Both consist of a subject concept combined with a predicate concept. As a result, someone could foolishly believe that the prime number between 3 and 5 is bald, and that "This sentence is false" is true. We could perhaps avoid these problems by adopting slightly vaguer language. We might say, with Frege, that a proposition is "something for which the question of truth or falsity can arise at all."[9] But then the next question must be why the question of truth or falsity arises for the thought "It will rain," but not for the thought "Will it rain?" The only answer possible, I believe, is that the former thought has an intrinsic, structural feature that the latter lacks. Alternatively, we might say that a proposition

9 Frege (1918: 5; 1919: 253) used this to define "thought," since he equated propositions with thoughts. Frege's formulation is also too broad in one respect, since it applies just as well to sentences as to the propositions or thoughts that they express.

is a thought with a structure capable of having true or false instances. The thought "This sentence is false" would count as a proposition because it has the same subject-predicate structure as "This computer is a Compaq," which is, in fact, true. While true, the idea that propositions are thoughts with a structure capable of having true or false instances cannot serve as a definition. For both interrogative thoughts and declarative thoughts are instances of more general structures. Such structures have true or false instances, but are too abstract to count as propositions.

The term "declarative" could be defined for thoughts within a Tarski-style theory of thought structure in several ways.[10] We could define a declarative thought as one that is not bound by the interrogative or imperative operators. Alternatively, we could define a declarative thought recursively as a subject-predicate thought, any disjunction or conjunction of declarative thoughts, and so on. While it is important for us to note that some but not all thoughts are propositions, we do not need to define the term "declarative." So we shall leave it primitive.

Definition 13.1 implies that propositions have the following familiar properties, because all are properties of thoughts. (1) Propositions are abstract entities. The proposition that the sky is blue does not have a spatial location, and exists eternally and necessarily (or timelessly and "worldlessly") in the way that numbers do. Propositions are abstract because they are *types*, specifically, types of mental events. Unlike "thought," "proposition" does not have a token sense in conventional usage. In no sense is there more than one proposition that the sky is blue. (2) Countless propositions have never been thought or believed. For while an event-token has necessarily occurred, many types of events have never occurred. There are no thought-tokens that no one has thought, but there are thought-types that no one has thought. If we use "thought" in its token sense, we would have to say that propositions are *actual or potential* thoughts.[11] But we are identifying propositions with thoughts in the type sense. Definition 13.1 does not identify propositions with *acts*

10 See my forthcoming *Nondescriptive Meaning and Reference*.

11 Cf. Leibniz 1677: 7, whose view is set forth in Church 1956b: 278–9 and Kretzmann 1967: 382. Bealer (1993b: 26–8) argues against identifying propositions with ideas on the grounds that propositions exist necessarily, whereas ideas are contingent, being mind-dependent. See also Brand 1984: 116. Of course, thoughts and ideas are mental entities; as such, they are vacuously mind-dependent. Moreover, thought and idea *tokens* are clearly contingent, because they are dependent on what happens to occur to actual thinkers. The causes and effects of thought are contingent on any view. But on the *ante rem* view we are adopting, the *existence* of thought and idea *types* is not similarly contingent (see §12.3).

of thought.[12] (3) Propositions are complex entities consisting of concepts and in some cases of other propositions or thoughts (see Chapter 14). (4) Propositions are expressed by, and therefore distinct from, sentences (see §13.1).[13] (5) Propositions are representations, with truth conditions and logical properties (see §23.3). (6) Propositions have their representational and logical properties essentially, in contrast to sentences. Whereas "The sky is blue" could have meant that grass is green, and would have if English had been different, the proposition that the sky is blue could not have failed to represent the sky as being blue. Hence the proposition is true iff the sky is blue, and has this truth condition essentially. (7) Propositions are intensions (extension determiners) composed of intensions. The extension (truth value) of a proposition is determined by the extension (denotation, reference) of its components. Different propositions may have the same extension and truth conditions. The proposition that seven is the square root of forty-nine is true if and only if the proposition that seven is the cube root of three hundred and forty-three is true. Nevertheless, the propositions are different, since one may be thought or believed while the other is not. Finally, (8) propositions are the relational objects of propositional attitudes.

The identification of propositions as a species of thoughts enables us to answer philosophical questions about propositional attitudes that are bound to worry empirical scientists: How can belief and desire affect our behavior and our emotions if they are relations to abstract objects? How can a relation to an abstract object have any effect on empirical reality? The answer is that these particular abstract objects are thoughts, which are a type of mental event. We stand in the belief relation to events of this type because of the causes and circumstances of their tokens, and their tokens have different effects when we believe them than when we do not.

§13.4 THE PROPOSITION THAT P

Let "p_i" be a declarative sentence expressing proposition i. Since propositions are thoughts (Definition 13.1) and thoughts are ideas (Definition 15.1), the generalization of Definition 7.5 tells us that we can refer to i as *the proposition* "p_i." The generalization of Definition 7.7 tells us that we can also refer to i as either *the thought that* p_i or *the proposition that* p_i

12 So we are not endorsing the "psychologism" that Husserl first advocated (1891) and then rejected (1900) under the force of Frege's critique.

13 Cf. Church 1956b: 276–7 and Cartwright 1962: 88–92.

(assuming that "p_i" is pronoun-free). All sentences of the following form are true.

13.2 **Theorem:** *The proposition (thought, idea) that* $p_i = i$*, provided i is a proposition.*[14]

This principle implies that sentences of form (2) (and cognate sentences with "thought" or "idea" in place of "proposition") are true in every context in which "the proposition expressed by 'p' " in its predicate is used to refer to the proposition expressed by "p" in its subject.

(2) The proposition that p = the proposition expressed by "p."[15]

As explained in §7.6, the singular terms flanking the identity sign in (2) are not synonymous. The expression on the right at most fixes the reference, not the sense, of the expression on the left. A speaker uttering "the proposition that grass is green" is not referring to a sentence. He *uses* the sentence "grass is green," but does not *mention* it. The speaker uses that sentence to express the thought that grass is green in the process of using the whole expression "the proposition that grass is green" to refer to the thought thus expressed.[16] The reference to the proposition is thus more demonstrative than descriptive. The function of the sentence is not to serve as an object of reference, but to introduce the proposition to be mentioned into the context. As a consequence, when we translate expressions of the form "the proposition that p" into another language, we translate "p" as well as "the proposition that." But when we translate "the proposition expressed by 'p,' " we do not translate "p" but leave it in English. Furthermore, "the proposition that grass is green" is a rigid designator, referring to the same proposition in every world, even worlds in which "Grass is green" does not express the proposition that grass is green. "The proposition expressed by 'Grass is green,'" by contrast,

14 Many treat that-clauses themselves as singular terms referring to propositions. I do not believe that-clauses function this way in standard English, and the oddity of the results often generates a spurious air of philosophical perplexity. Witness Schiffer: "What . . . is the nature of this thing, *that eating fish increases intelligence*, which is the referent of the that-clause singular term? Well, that eating fish increases intelligence is *abstract* in that it has no spatial location. . . . " (1994: 280). What is true is that "that p" sometimes appears as an optional variant in contexts in which "the proposition that p" occurs, which is a genuine singular term.

15 Cf. Richard 1981: 3.

16 This differs markedly from Frege's view, which maintains that the sentence in a that-clause has an oblique sense in which it refers to the proposition that is its customary sense (see §21.2).

is a nonrigid designator, and does not express the proposition that grass is green in a world in which "Grass is green" means that snow is purple. Sentences of form (2) are thus like "I am the person uttering 'I,'" which is true in every context in which "I" is used deictically, but does not express a necessary truth in any context.

Even though "the proposition that p" and "the thought that p" refer to the same object, they are not synonymous, and are not used interchangeably. We have already observed that "thought" is more general than "proposition." That is, there are objects of thought other than propositions. And even when both terms apply, we generally use the term "proposition" when we wish to stress the logical characteristics of the thoughts (truth, probability, entailment, etc.) and abstract away from their ontological characteristics (the fact that they are types of mental events that occur to various people at various times and have various causes and effects). We use the term "thought" with the opposite stress. Thus "thinking the thought that . . . " is more idiomatic than "thinking the proposition that. . . . " And "The thought that p implies . . . " usually means "Thinking that thought implies . . . " not "The truth of that thought implies. . . . " "The proposition that p implies . . . " always means the latter, never the former.[17]

One argument against the identification of propositions with thoughts goes as follows. In the object sense of "belief," the belief that p is the proposition that p. Yet it would be odd to say that the belief that p is the thought that p, especially when we have emphasized that speakers often think and express the thought without believing or expressing the belief. The claim that the belief is a thought seems to imply that believing is thinking, which we have taken pains to deny. This cannot be a sound objection, however. For clearly, what S is thinking may be identical to what S believes; and in the relevant sense, what S is thinking is a thought, and what S believes is a belief. One way of accounting for the divergent linguistic intuitions is to hypothesize that in English, "the belief that p" means something like "the proposition that p *qua object of belief*," that "the thought that p" means "the proposition that p *qua object of thought*," and so on for desire, intention, hope, knowledge, statement, prediction, and so on. "The proposition that p" itself may be taken to mean simply "the declarative idea that p." Then we may note that the oddity of "George

17 While the truth of the thought that today is hot does follow logically from the truth of the thought that today is hot and humid, it would be misleading to say that the former thought logically implies the latter. For that would ordinarily be interpreted as meaning that *thinking* the former thought logically implies thinking the latter, which is a very different claim.

Bush *qua* father is George Bush *qua* president" does not show that "George Bush *qua* father" and "George Bush *qua* president" do not refer to one and the same person, George Bush. Similarly, from the oddity of "The noun *cut* is the verb *cut*," it does not necessarily follow that "The noun *cut*" and "The verb *cut*" do not both refer to the single word *cut*, which may appear either as a noun or as a verb. It seems that a phrase like "the noun *cut*" means something like "the word *cut* as a noun." In official formulations, we will use the singular terms introduced by Theorem 13.2 purely referentially. Any "qua" implications will be treated as connotations rather than as parts of the denotation.

My account of that-clauses is similar to Davidson's (1968) in key respects. Davidson focuses on "saying that," and suggests that "Galileo said that the Earth moves" can be viewed as equivalent to "The Earth moves. Galileo said that." Davidson makes it clear that the speaker is not referring to the sentence "The Earth moves," but is using it to refer to the Earth and to represent it as moving.

> The proposal then is this: sentences in indirect discourse . . . consist of an expression referring to a speaker, the two-place predicate "said," and a demonstrative referring to an utterance. Period. What follows gives the content of the subject's saying, but has no logical or semantic connection with the original attribution of a saying. . . . "Jones asserted that Entebbe is equatorial" would, if we parallel the analysis of indirect discourse, come to mean something like, "An utterance of Jones' in the assertive mode had the content of this utterance of mine. Entebbe is equatorial." (Davidson 1968: 344–5)[18]

Despite the important similarities, many parts of Davidson's proposal are literally false. The "that" in "the proposition that p" or "S said that p" is a subordinating conjunction, not a demonstrative pronoun. The speaker does not *refer to* his utterance of "p," and does not actually *say that p*. A speaker who says "Galileo said that the Earth moves" has said that Galileo said something, but has not said that the Earth moves. Finally, "p" obviously does have a "semantic" connection to "the proposition that p" and to "S said that p." For the meaning of these compounds depends compositionally on the meaning of "p," in the normal way (Chapter 10). It might be thought that Davidson avoided reference to "intensional entities" like propositions. But Davidson makes it clear that that-clauses refer to the *content* of the speaker's utterance, not merely to his act of uttering something.

18 Cf. Peacocke 1975: §4 and Heal 1997: 638–9. See especially Lycan (1981; 1985) and Böer and Lycan (1986: Chapters 3–4), who eliminate most of the defects in Davidson's analysis.

My account of that-clauses is also similar to Forbes's (1990, 1993, 1997b, 1997c). Expressions of the form "the proposition that p" typically, although not exclusively, occur as the grammatical object of propositional attitude verbs like "believes." Forbes takes a sentence like "S believes the proposition that p" to assert that S stands in a certain relation to the *"so-labeled"* proposition, where "so-labeled" is a "logophor" referring to the sentence "p." Forbes takes propositions to be complexes of senses, which are expressed by words. He says that a word "labels" a sense when it expresses that sense. I would object here that "label" sounds too much like "names" (contrast §21.2) and that propositions cannot be identified with senses (§21.1). But these objections are immaterial in the present context, since everything can be done in terms of "proposition" and "express." When we do this, we get the claim that "S believes the proposition that p" asserts that S stands in a certain relation to the proposition expressed by "p." Once this is modified to account for the fact that when "p" is ambiguous, it may be used to express different propositions on different occasions, we get the claim that (2) is true in any context in which "the proposition expressed by 'p' " in its predicate is used to refer to the proposition expressed by "p" in its subject.

Forbes, however, does not take "the so-labeled proposition" to refer to the proposition that "p" expresses on the occasion in which "S believes the proposition that p" is used, which is determined by the conjunction of what the speaker means by "p" and what "p" means in English (or, more generally, in the metalanguage that the speaker is using). Instead, he takes "the so-labeled proposition" to refer to *the proposition S would express using his linguistic counterpart of "p"* (1993: 315). Forbes says little about what the subject S's linguistic counterpart of a sentence is. His most explicit remarks make it appear that S's linguistic counterpart of "p" is a sentence in the subject's language that has the same meaning as "p" in the speaker's language (see, e.g., Forbes 1990: 551–2). Even on this assumption, there are several problems with Forbes's suggestion. First, there is the possibility that S would misspeak, using her linguistic counterpart of "p" to express a proposition that is not expressed by "p" in her language. Second, there is the possibility that S would use her linguistic counterpart of "p" to express a proposition expressed by "p" in her language that is not the proposition meant by "p" when the speaker said "S believes the proposition that p." (S's linguistic counterpart of "p" may have meanings in S's language that "p" does not have in English.) In either case, the proposition that S would express using her linguistic counterpart of "p" is not the proposition that p. These difficulties can be

avoided by taking "the so-labeled proposition" to refer to *the proposition expressed by a sentence in S's language meaning what "p" means in "S believes that p" on this occasion*. But this is just a long-winded way of saying that "the so-labeled proposition" refers to *the proposition expressed by "p" in "S believes that p" on that occasion*.

The fact that Forbes takes "the so-labeled proposition" to refer to a proposition that the *subject* of the belief attribution would express using *her* language rather than to the proposition the *speaker* expresses as part of *his* belief attribution creates problems for belief attributions to non-linguistic animals, and for references to propositions outside the context of propositional attitudes. If I say "Samantha believes that food is in the dish," my statement would not be false or truth-valueless simply because Samantha is a cat and does not have a language. And when a speaker says "The proposition that Mary is a vixen is true, and entails the proposition that Mary is a woman," the speaker is not ascribing a propositional attitude to any subject. So what anyone other than the speaker would express is irrelevant. Without a uniform treatment of what descriptions of the form "the proposition that p" refer to, a Forbesian account will not be able to secure the validity of inferences like "S believes the proposition that p; the proposition that p is true; therefore S believes something that is true."

Why does Forbes reject an "ascriber-oriented" view of that-clauses in favor of his "believer-oriented" view? First, he thinks that the speaker uses the words he does because he heard the believer use them, or believes that the believer "associates" those words with certain "dossiers" (Forbes 1990: 549). This may be true in some cases, but cannot be true when the speaker is ascribing beliefs to nonlinguistic animals, or using a language or code that he knows the believer does not know. Second, Forbes offers the following argument.

> The ascriber-oriented view is refuted by the case of Ralph, for whom "Superman" and "Clark Kent" express the same sense. There is a clear intuition that Ralph can utter ["Lois believes that Superman can fly and Clark Kent can't"] and say something true; yet the propositions *Ralph* expresses by "Superman can fly" and "Clark Kent can't" are explicitly contradictory, though Lois does not believe an explicit contradiction. So Ralph is not referring to his own propositions. (Forbes 1990: 546)

If Ralph is using "Superman" and "Clark Kent" not just as co-referential terms, but as *having the same sense*, then he is not using these terms with their conventional English senses. Hence it is irrelevant that when *Ralph* uses "Lois believes that Superman can fly and Clark Kent can't"

he expresses a proposition that ascribes a contradiction to Lois. For we are trying to figure out what "S believes that p" means as conventionally used in English (and what its translations mean in other languages). The proposition expressed by "p" on a given occasion is partly determined by what the speaker means by it. But it is also determined by what "p" means in English (see §7.9).

§13.5 SITUATIONS AND POSSIBLE WORLDS

I have defined a proposition as a declarative thought, and have asserted as a substantive claim that other propositional attitudes can be analyzed as relations to propositions in this sense. I believe that this definition fits the usage of "proposition" in conventional English, and captures one of its standard uses in philosophy, logic, and linguistics. But as noted earlier, "proposition" is a technical term with many other common definitions. On one, propositions are declarative sentences. Since propositional attitudes are not relations to propositions in this sense (§13.2), this definition is not useful for our purposes.

In a situation semantics, propositions are defined as situations, and propositional attitudes are characterized as relations to situations.[19] By *"situation"* (or *"state of affairs"*), I mean an ordered n-tuple consisting of n objects plus an n-adic property or relation. To believe that Bill Clinton is a Democrat, on this approach, is to stand in a relationship to Clinton and the property of being a Democrat. The major difficulties for the situation semanticist's theory of propositional attitudes have come to be known as *Frege's* and *Russell's problems*. Frege's problem arises from the intensionality of propositional attitude contexts, the fact that substitution of coextensive terms in such contexts does not always produce equivalent statements. Someone can believe that Cary Grant is famous, for example, without believing that Archibald Leach is.[20] This would not be possible if belief were a dyadic relation between believers and

19 Cf. Kaplan 1977; Perry 1977; 1979; 1980; 1990; 1997a; 1997b; Schiffer 1978; 1987b; 1990; 1992; 1995; Böer & Lycan 1980; McKay 1981; Barwise & Perry 1981a; 1981b; 1983; Salmon 1981; 1986: Chapters 8–9; 1989b; 1990; Stich 1983: 48–9; Fitch 1985; 1987; 1990; Soames 1987a; 1987b; Richard 1987; 1989; Crimmins & Perry 1989; Austin 1990: 112–15; Fodor 1990a; 1994: 39–49; 1998a: 12–17; Braun 1991; 1993; Crimmins 1992; Recanati 1993; Adams, Stecker, & Fuller 1993a; 1993b; Adams & Stecker 1994; Adams, Fuller, & Stecker 1997; Pelczar & Rainsbury 1998. Contrast Forbes 1989: 474–5.

20 Remember that I am interpreting belief descriptions opaquely rather than transparently (§6.2). Interpreted transparently, "S believes Archibald Leach is famous" will be true if S believes Cary Grant is famous.

situations. Since Cary Grant is Archibald Leach, the n-tuple consisting of Cary Grant and being famous is identical to the n-tuple consisting of Archibald Leach and being famous. Anyone related in any way to the first situation must be related in the same way to the second. Russell's problem arises from the intentionality of propositional attitudes, the fact that people can think about nonexistent objects and have other propositional attitudes concerning them. Many children believe that Santa Claus brings presents at Christmas, even though Santa Claus does not exist. This would not be possible if propositional attitudes were relations to situations. For an n-tuple exists only if its elements exist. Since Santa Claus does not exist, there is no n-tuple consisting of Santa Claus and the property of bringing presents. There is no such thing as the proposition that Santa Claus brings presents if the proposition that s is P is the n-tuple consisting of the referent of "s" and the property expressed by "is P."

The situation semanticist commonly responds to Frege's problem by abandoning the idea that propositional attitudes are dyadic relations between individuals and situations. Instead, propositional attitudes are taken to be triadic relations linking individuals, situations, and "*modes of presentation*" or "*ways of believing.*" On this approach, "S believes that Cary Grant is famous" and "S believes that Archibald Leach is famous" are taken to describe S as being related to the same situation in different ways, which involve different modes of presentation of the single individual who is both Cary Grant and Archibald Leach. This solution immediately raises the question of what these modes of presentation are. The most promising answer, I believe, is that they are propositions or thoughts in the sense that I have been clarifying – that is, mental representations consisting of complexes of concepts. Since the concept of Cary Grant is not the same as the concept of Archibald Leach, they are different modes of presentation of the actor. What the situation semanticist calls a proposition, I would call the extension of a proposition. As far as Frege's problem goes, situation semantics and ideational semantics differ only in terminology. But the mode-of-presentation gambit is no help with Russell's problem. Since Santa Claus does not exist, there is no ordered n-tuple containing Santa Claus for people to believe. So believing that Santa Claus exists cannot be treated as any kind of relation to a situation.

In a possible-worlds semantics, propositional attitudes are characterized as relations between individuals and sets of possible worlds (or equivalently, as functions from possible worlds to truth values), and propositions

are defined as world-sets.[21] To believe that p is to be related to the set of all possible worlds in which it is true that p. Frege's and Russell's problems still arise, because even the substitution of necessarily equivalent propositions is not guaranteed to preserve the truth value of a belief statement, and because people can have beliefs about impossible objects. People can believe that 7 is the square root of 49 without believing that 7 is the cube root of 343, even though "7 is the square root of 49" and "7 is the cube root of 343" are true in the same set of possible worlds. And someone might believe that the smallest positive real number is odd even though there is no possible world containing a smallest positive real number. Any useful results of possible-worlds semantics can be obtained without the cost of denying such familiar possibilities. For the world-sets that the possible-worlds semanticist calls propositions are the *intensions* of the thoughts that I call propositions. In virtue of believing propositions in my sense, the believer stands in a distinctive relationship to intensions, if they exist. The problems arise only when belief is identified with the relation between believer and intension rather than that between believer and thought.[22]

§13.6 SEMANTIC THEOREMS AND THE MATES OBJECTION

If "p" expresses a proposition, then it means something, namely "p" (Definition 7.4). So if "p" is meaningless, it cannot express a proposition. Hence there is no proposition that p. This is the well-known *negative existence criterion* for propositions.

13.3 **Theorem:** *If "p" is meaningless, then it expresses no proposition.*

Thus "Gorky joks nedzik" expresses no proposition in English (or in any other language, to my knowledge). Hence there is no such thing as the proposition or thought that gorky joks nedzik. The expression "the proposition that gorky joks nedzik" fails to refer to anything because the sentence forming its that-clause does not express any proposition (Theorem 13.2).

As "The square root of four equals two" and "$\sqrt{4} = 2$" illustrate, two sentences expressing propositions express the same proposition *if* they

21 See Hintikka 1963; 1969; Stalnaker 1976a; 1976b; Field 1978; Lewis 1979a; 1981b; 1986; Cresswell 1985.

22 I explore situation and possible worlds semantics more deeply in *Nondescriptive Meaning and Reference*.

have the same meaning. It is also the case that sentences express the same proposition *only if* they have the same meaning. Both results follow from Definition 7.4 (relating meaning to the idea expressed) together with Definition 15.1 (saying that thoughts are ideas) and Definition 13.1 (saying that propositions are thoughts). So we can formulate the following *identity criterion* for propositions.

13.4 **Theorem:** *If "p" and "q" express propositions P and Q respectively, then P = Q iff "p" and "q" have the same meaning.*

Of course, we mean having *exactly* the same meaning. Identity of propositions is correlated with identity of meaning; similarity of meaning indicates only similarity of proposition.[23] This criterion is appropriate on our definition of a proposition as a declarative thought, but not on the others mentioned earlier. For example, "Bill's father is sick" and "Jane's brother is sick" clearly differ in meaning, even if Bill's father happens to be identical to Jane's brother. And in accordance with Theorem 13.4, the sentences express different propositions: the thought that Bill's brother is sick can occur to us without the thought that Jane's brother is sick, and vice versa. Nevertheless, if Bill's father happens to be identical to Jane's brother, the two sentences represent the same state of affairs – the same object's being sick. Similarly, Theorem 13.4 rules that "7 is the square root of 49" and "7 is the cube root of 343" express different propositions because the sentences differ in meaning, despite the fact that they are true in the same set of possible worlds.

It follows from Theorems 13.2 and 13.4 that *the proposition that p is identical to the proposition that q if and only if "p" and "q" are synonymous* (assuming that "p" and "q" express propositions). Hence the proposition that the square root of 4 equals 2 and the proposition that $\sqrt{4} = 2$ are the same proposition. But the proposition that $\sqrt{4} = 2$ is very different from the proposition that $\sqrt{2} = 4$, because "$\sqrt{4} = 2$" and "$\sqrt{2} = 4$" have different meanings.

The semantic theorems can be generalized to cover all thoughts, not just propositional thoughts. First, *a sentence expresses a thought only if it is meaningful*. "Gorky joks nedzik" does not express an interrogative, imperative, or declarative thought. Second, *meaningful sentences express the same thought iff they have the same meaning, assuming that they express thoughts.* "Mary will sing," "Will Mary sing?" and "Mary, sing!" express different

23 Thus Quine's (1960: 201–3) objection that synonymy but not identity is graded can be avoided. Quine's general concerns about synonymy will be addressed in §21.5.

thoughts, since they have different meanings. "Mary, sing!" and "Sing, Mary!" express the same thought, because they have the same meaning.

Sentences with indexicals may appear to provide counterexamples to Theorem 13.4. Thus sentences (3) and (4) can be used to express the same proposition, but do not mean the same thing.

(3) Mary kissed Jack.
(4) Mary kissed him.

However, Theorem 13.4 correlates propositional identity with identity of *sentence* meaning, not *speaker* meaning. While speakers can use (3) and (4) to mean the same thing, the sentences themselves do not mean the same thing. Whereas (3) expresses a definite nonindexical proposition in English, (4) does not. The meaning of (4) allows it to be used to express a broad range of nonindexical propositions with a common structure, but no one of those propositions can be singled out as what (4) expresses. It cannot even be said that (4) is ambiguous, expressing all of those propositions. For the sentence is not used in a different sense when it is used to make a statement about Tom, Dick, or Harry, rather than about Jack. Sentence (4) itself expresses either an indexical proposition, or none at all. In either case, it does not express the same proposition as (3), so there is no violation of Theorem 13.4.

In a discussion of the location of the tallest mountain, Mt. Everest, it would be natural to claim that (5) and (6) "say the same thing."

(5) Mt. Everest is in Nepal.
(6) The tallest mountain is in Nepal.

Nevertheless, the sentences do not have the same meaning, since "Mt. Everest" and "the tallest mountain" are not synonyms. Again, there appears to be a violation of Theorem 13.4. However, the two sentences do not express the same proposition in the sense defined by Definition 13.1. "Mt. Everest is in Nepal, but the tallest mountain is not" is not a self-contradiction; it is false, but not logically false. Furthermore, someone who believes the proposition that the tallest mountain is in Nepal may never even have considered the proposition that Mt. Everest is in Nepal. Since there is a difference both in meaning and in the proposition expressed, there is no violation of Theorem 13.4. "Saying the same thing" does not generally require identity of meaning or of the proposition expressed. It does so only when the purposes of the context require extremely fine discriminations of what is said, as in discussions of the logical

validity of arguments. In a logic course, when observing that the argument from (5) to (6) is deductively invalid, it would be most unnatural to insist that the premise says the same thing as the conclusion. For that would suggest that the argument is question-begging and therefore valid.

The main objection to Theorem 13.4 is based on a form of argument developed by Mates (1950).[24] It seems evident, given the meaning of (exact) synonymy, that except in idioms and quotations, substituting one synonym for another in a larger expression (and making no other changes) should result in a synonymous and therefore equivalent expression. This is called the *Principle of the Substitutivity of Synonyms*, or the substitutivity principle for short. Quotation is an exception to substitutivity, because words do not have their standard meanings when they are being used to refer to themselves. Idioms are an exception because their meanings are not determined compositionally by the meanings of their components. The Mates objection concludes, however, that substitution of synonyms in the complements of psychological verbs also fails to produce synonyms. For example, "undertaker" and "mortician" are commonly given as synonyms, as are "oculist" and "eye doctor," and "brother" and "male sibling" (see, e.g., *Webster's Encyclopedic Unabridged Dictionary*). If these terms are synonymous, then the most elementary form of compositionality implies that (7) and (8) are synonymous.

(7) Undertakers are undertakers.
(8) Undertakers are morticians.

If synonyms are interchangeable in propositional attitude contexts, then (9) should be synonymous with (10), and so on for higher-order iterations and other propositional attitudes.

(9) Henry believes that undertakers are undertakers.
(10) Henry believes that undertakers are morticians.

24 The Mates objection is a simple reformulation of Moore's (1942) "paradox of analysis." Like Moore, Mates himself accepted Theorem 13.4 and used it to point out an odd consequence of Carnap's (1947: §§13–15) analysis of synonymy as intensional isomorphism. Putnam (1954) inverted the argument, assuming some synonyms and using them to reject Theorem 13.4. It is Putnam's inverted formulation that is called "the Mates Objection." See also Scheffler 1955; Quine 1960: 202–3; Burge 1978; Linsky 1983: 33–9; Boër 1986: 93; Loar 1987: 180; Fodor 1989: 66–8; 1998a: 16–17; Richard 1989: 328; Tomberlin 1991: 241; Crimmins 1992: 26; Nelson 1992; Bealer 1993a: 19; Taschek 1995: 73–4. My colleague Linda Wetzel raised the objection in a private correspondence, using "actor" and "thespian." Contrast Church 1954; Carnap 1954; Sellars 1955; Pap 1955; 1957; Lakoff 1971a: 282; Katz 1972: 267–74; Yagisawa 1984; Soames 1987a: 123–5; Crane 1991: §6; Fodor 1994: 112; Horwich 1998a: §3.24.

Theorem 13.4 alone does not entail that (9) and (10) are *synonymous*. But together with Definition 13.2 and the assumption that "believes" is a relational term, Theorem 13.4 does entail that (9) and (10) are *equivalent* if "undertaker" and "mortician" are (exact) synonyms. For in that case, the proposition that all undertakers are morticians *is* the proposition that all undertakers are undertakers.

13.5 **Theorem:** *If $i = j$, then S believes the proposition that p_i iff S believes the proposition that q_i.*

That is, "S believes the proposition that p" is true iff "S believes the proposition that q" is true when "p" and "q" are synonymous. "S believes the proposition that p" can be reduced here, as elsewhere, to "S believes that p" and further to "S believes p" *salve veritate.* Theorem 13.5 can be generalized to thought and to all the other propositional attitudes. To derive the stronger theorem that the two belief sentences are synonymous when "p" and "q" are synonymous would require the linguistic principle that the meaning of "S believes that p" is determined compositionally by the meanings of "S," "believes," "that," and "p." This principle seems evident, but it is not one of our theorems. The principle of compositionality for belief sentences provides independent support for Theorem 13.5.[25]

The Mates objection maintains that (9) and (10) are neither synonymous nor equivalent. It is easy to imagine Henry saying "Not all undertakers are morticians," or "I don't know whether undertakers are morticians," and meaning it. This would be evidence that (10) is false, but it is no evidence at all that (9) is false. Similarly, it is easy to imagine someone believing that (9) is true while denying (10). Indeed, I believe that (11) would be true in the case imagined, while (12) appears to be false.

(11) Putnam believes that Henry believes that undertakers are undertakers.
(12) Putnam believes that Henry believes that undertakers are morticians.

But if (11) and (12) have different truth values, they must have different meanings. That would entail that (9) and (10) and therefore (7) and (8) have different meanings if synonyms were interchangeable in propositional attitude contexts. If the substitutivity principle were correct for all propositional attitudes, the fact that anyone had even the slightest *doubt* that whoever believes d believes d′ would entail that "d" and "d′" are

25 It should be noted that in accepting substitution of synonyms for synonyms in that-clauses *salva veritate*, a principle justly associated with Frege, I have *not* offered Frege's explanation. See §21.2.

not synonymous. Since Putnam thought that some doubt would always be possible as long as "d" and "d′" are different sentences, he reasoned that the substitutivity principle would imply that there are no synonyms whatsoever, which he took to be a reductio ad absurdum. The Mates objection maintains, that is, that even though a pair of sentences like (7) and (8) are synonymous, it is possible for someone to doubt that whoever believes (7) believes (8). Hence the substitutivity principle must be false.

The standard reply to the Mates objection is that Henry's utterance of "Not all undertakers are morticians" is evidence that Henry does not realize that "undertaker" and "mortician" are synonyms; but if Henry does not properly interpret the sentence he has uttered, then we cannot infer that he believes the proposition that it actually expresses. So if we believe that "undertaker" and "mortician" are synonymous, then we cannot infer that (10) is false from Henry's utterance. However, we cannot give the same explanation for the apparent difference between (11) and (12) without begging the question. The Mates objection assumes that "undertaker" and "mortician" are synonymous and argues that (9) and (10) are not, drawing the general conclusion that substitution of synonyms need not produce a synonymous belief sentence. It would therefore beg the question to take Putnam's affirmation of (10) and denial of (9) as evidence that Putnam mistakenly assumes that (9) and (10) are not synonymous.

Church's (1954) reply is more effective. If "mortician" and "undertaker" are exact synonyms and are purely used, then we should be able to produce an exact translation by rendering both using the same word in another language. We may have to use the same word if the language does not have synonyms paralleling those in English. But then the translations of (11) and (12) would be identical, making it impossible for (11) and (12) to differ in truth value. However, it does seem correct to deny (12). "Putnam believes the proposition that Henry doesn't believe all undertakers are morticians" appears to be true. So it would seem that any translation rendering (11) and (12) identical must be imperfect.

I submit that the Mates objection is inconclusive for two reasons. First, (12) may appear to be false only because "undertaker" and "mortician" are being *mentioned* rather than *purely used*. It would be easy for Henry to believe that all undertakers are *called "undertakers"* without believing that they are *called "morticians."* But this is irrelevant to Theorem 13.5. For substitution of synonyms inside quotation marks obviously need not produce synonyms. We often use object-language formulations in everyday English when we intend our claim to be metalinguistic. Thus if someone

uses a German word that we do not understand, and we say "I do not know what a Vogel is," it would be inappropriate or facetious to reply "Of course you do, you see one [a bird] every day." Similarly, if a foreign medical resident unsure of his English is asked to give his diagnosis and pronounces a "perforated ventricle," he may turn to his friend and say "At least I hope it is a perforated ventricle," meaning not that that is the condition he wishes on his patient, nor that he is unsure of the nature of the patient's condition, but that he hopes he has used the right terminology. Finally, the familiar "My name is John" does not claim that my name is a person (Yagisawa 1984: 416).

I think that when "G" is a relatively uncommon synonym of a common term "F," or is the word for "F" in another language, "S does not know that Fs are Gs" is normally intended metalinguistically. As Sellars (1955) astutely observed, such a denial of knowledge would be so implausible if the speaker were purely using the terms "F" and "G," and using them as synonyms, that we naturally adopt the more charitable interpretation, on which the speaker is mentioning the words. Thus "Tom did not know that an electroencephalograph is an EEG" would normally be interpreted as saying that Tom did not know that an electroencephalograph is called an "EEG," not that Tom did not know that an electroencephalograph is an electroencephalograph. One way to force ourselves to treat "undertaker" and "mortician" as stylistic variants is to insert a parenthetical "i.e., undertaker" after "mortician." When we do this, it is hard to hear (12) as saying something false, unless we take it to claim that Putnam thinks Henry is mentally defective. Thus (10) may seem to be false because "Henry does not know that undertakers are morticians" is true metalinguistically. But Theorem 13.4 rules that (9) and (10) are equivalent only when they are interpreted literally.

Second, it is not *clear* that "undertaker" and "mortician" *are* synonyms. To my ear, "mortician" seems to denote a subclass of undertakers, those for whom undertaking is a profession. I can imagine someone being an untrained, temporary, volunteer undertaker during times of civil disaster, or someone who is an undertaker in his spare time for fun. I do not think that would make him a mortician. "Professional mortician" seems redundant, "professional undertaker" does not. I believe that "mortician" stands to "undertaker" somewhat as "beautician" stands to "hairdresser" and "physician" stands to "healer." I may be wrong about this. As I said, I find the case unclear. The question can be settled only by detailed empirical study of the usage of competent speakers of English. The fact that my dictionary defines a mortician simply as an undertaker is certainly

evidence that they are synonyms. But dictionary entries are often not even intended to be synonymous, and are sometimes inaccurate. See, for example, the entries in *Webster's Encyclopedic Unabridged Dictionary* for "gold" ("a precious yellow metallic element, highly malleable and ductile, and free from liability to rust. *Symbol*: Au; *at. wt.* 196.967; *at. no.* 79; *sp. gr.*: 19.3 at 20°") and "desire" ("a longing or craving"). The former entry is a concise, scientifically accurate characterization of the reference of "gold," but does not, and presumably was not intended to, have the same sense as "gold." The latter is obviously too narrow as a definition. Not all desires are strong enough to be cravings, or are directed at things absent.

The hypothesis that "undertaker" and "mortician" are not synonyms provides a simple explanation for the fact that competent speakers like Putnam might believe that (10) could express a falsehood even though (9) is true, an explanation that does not require a large class of exceptions to the principle of compositionality. Similar evidence casts doubt on the synonymy of "oculist" and "eye doctor." The fact that "sibling" is learned later and less frequently than "brother" weighs against the synonymy of "brother" and "male sibling." The fact that languages have fewer exact synonyms than we might previously have thought is no more earthshaking than the fact that fewer terms have analytic definitions than has traditionally been believed.[26]

"Synonymous" is sometimes used strictly to denote *exact* sameness of meaning. In this usage, "premise" and "premiss" count as synonymous, but not "desire" and "crave," which are commonly listed as synonyms in dictionaries and synonymies. But "synonymous" is also used more generally, to denote *approximate* sameness of meaning, where what counts as a good approximation varies from context to context. In this broader sense, "desire" and "crave" are synonyms. In this respect, "synonymous" resembles "flat" and "straight," which are sometimes used in an exact sense (as in geometry) but more often in a variably strict sense (as in application to roads). "The same" is intended to have its strict sense in Theorem 13.4. Hence the Mates example constitutes a counterexample to Theorem 13.4 only if "mortician" and "undertaker" are exact synonyms. It is obvious that "mortician" and "undertaker" are synonymous in the loose sense. It is not clear that they are exactly synonymous. I have presented evidence that they are not, while conceding that the evidence is not conclusive.

26 See also §8.5, §15.2, and *Nondescriptive Meaning and Reference*.

Burge (1978) assumed that (a) "fortnight" and "fourteen-day period" are synonyms, while presenting evidence that (b) speakers can use "A fortnight is a ten-day period" to express a belief other than the obvious falsehood that a fourteen-day period is a ten-day period, and that (c) speakers can use "S believes that a fortnight is a ten-day period" to attribute to S a belief other than the belief that a fourteen-day period is a ten-day period. Burge is concerned here with what the rules and conventions of English allow the speaker to express. He is not making the trivial Humpty-Dumpty point that speakers can depart from conventional usage if they want to. From these premises considered jointly – and others like (b) and (c) – Burge concluded that synonyms are not substitutable in belief clauses *salve veritate*. Before we can accept this argument, however, we have to know that Burge was right in assuming that "fortnight" and "fourteen-day period" are synonyms. The fact that a fortnight is defined as a fourteen-day period in dictionaries supports Burge's assumption, but it is not decisive, as we have seen. The minute we raise the question of whether the terms have exactly the same sense, we realize that the linguistic evidence represented by (b) and (c) undermines (a), and vice versa. We cannot maintain that "fortnight" and "fourteen-day period" are (exact) synonyms if speakers can use "A fortnight is a ten-day period" to express a thought or belief other than the obvious falsehood that a fourteen-day period is a ten-day period. The conclusion that "water" and "H_2O" do not have the same meaning despite having the same reference is supported in the same way by the fact that competent speakers can properly use "Water is not H_2O" to express a consistent, empirical proposition that some people (like John Dalton) have mistakenly thought was true, and that many people (like young children, and everyone who lived before atomic theory was developed) could not even conceive. The interconnectedness of our judgments of synonymy and attributions of belief is powerful support for the semantic theorems.

I believe that clear cases confirm Theorem 13.4. Consider "premises" and "premisses," "five" and "5," or "Bill's accuser" and "The accuser of Bill." These pairs are clearly exact synonyms. Moreover, hardly anyone knows one but not the other or uses them as anything but stylistic or unconscious variants. When these synonyms are substituted in belief contexts, the results are semantically indistinguishable.

(13) Henry believes that premises are premises.
(14) Henry believes that premises are premisses.

I can detect no difference in meaning between these sentences.[27] If Henry writes "Premises are somewhat questionable premisses, so not all premises are premisses," we could only conclude that Henry does not know what he is talking about, and is ignorant of one fact about English. If Henry insists that *he* means "questionable premiss" by "premise," we need not object. Indeed, we would properly conclude that *he* uses "all premises are premisses" and (14) to express propositions that are false. And that conclusion would accord with Theorem 13.4, when relativized to S's idiolect. But we are using standard English. It is in English that "premises" and "premisses" are synonymous. And in English, "all premises are premisses" and (14) express propositions that are true. For these are exactly synonymous with "All premises are premises" and (13), respectively, which clearly express true propositions. It is hard to imagine what could constitute good evidence that (14) is false while (13) is true. But if there were such evidence, it would *also* be evidence that "All premises are premises" and "All premises are premisses" are not in fact synonymous. Even though his usage of the terms may be evidence for its truth value, it cannot be maintained that the truth value of (14) depends on Henry's usage or understanding of the terms "premises" and "premisses." For (14) could be true even though Henry is a Mongolian who does not know a word of English.

Salmon presents an interesting variant of the Mates problem, in an effort to show that even the clearest synonyms are not substitutable.

> Suppose that foreign-born Sasha learns the words 'ketchup' and 'catsup' not by being taught that they are perfect synonyms, but by actually consuming the condiment and reading the labels on the bottles. Suppose further that, in Sasha's idiosyncratic experience, people typically have the condiment called 'catsup' with their eggs and hash browns at breakfast, whereas they routinely have the condiment called 'ketchup' with their hamburgers at lunch. This naturally leads Sasha to conclude, erroneously, that ketchup and catsup are different condiments that happen to share a similar taste, color, consistency, and name. He thinks [says?] to himself, 'Ketchup is a sandwich condiment, but no one in his right mind would eat a sandwich condiment with eggs at breakfast; so catsup is not a sandwich condiment.' Whereas the sentence 'Ketchup is ketchup' is uninformative for Sasha, the sentence 'Catsup is ketchup' is every bit as informative as 'Hesperus is Phosphorus'. Applying the general strategy invoked in Frege's classic argument against Millianism, we would conclude that the terms

27 If you think this is because "premise" and "premiss" are pronounced the same, change to a spoken example, say "ehconomics" and "eeconomics" or "toppin" and "topping."

'catsup' and 'ketchup' differ in information value for Sasha. But this is clearly wrong. The terms 'ketchup' and 'catsup' are perfect synonyms in English. . . . Each word (spelling) was learned by Sasha in much the same way that some of us learned it. Even in Sasha's idiolect, then, the two words (spellings) are perfectly synonymous, and therefore share the same information value. (Salmon 1989b: 216–17)

The fact that each word was learned by Sasha "in much the same way that some of us learned it" is *some* evidence that the two words have the same meaning for Sasha that they have for us. But *pace* Wittgensteinian forms of behaviorism, the evidence is not conclusive. Indeed, "ketchup" and "catsup" cannot have the same meaning in Sasha's idiolect. For by stipulation, Sasha uses "Ketchup is a sandwich condiment" and "Catsup is a sandwich condiment" to express *different* thoughts, and Sasha uses "Ketchup is not catsup" to express a contingent proposition that he takes to be true. I flagged Salmon's use of "thinks" in the passage quoted because given that "catsup" and "ketchup" are exact synonyms in English, which is the language *we* are using, *we* cannot say that Sasha *thinks* "Ketchup is a sandwich condiment, but catsup is not" unless we are willing to attribute a self-contradictory thought to Sasha. We can, of course, allow that Sasha *says* this (to himself or out loud) without attributing to him a contradiction. But that is only because he might mean something different by his terms that we do. Note finally that when Salmon says "This naturally leads Sasha to conclude, erroneously, that ketchup and catsup are different condiments," we are forced to interpret the claim metalinguistically. For if *Salmon* were using "ketchup" and "catsup" to mean *exactly the same thing*, he would be making an astonishing claim about Sasha. Imagine translating the claim into a language that only has one word for the tomato-based condiment.

It might be insisted that even the examples of clear synonyms I have given are debatable, there being no exact synonyms in English or any other natural language. While I think that such a view is implausible in the extreme, we could still get confirmation of Theorem 13.4 by *stipulating* our synonyms. Let us invent a code in which names of aircraft are replaced by names of familiar animals. Let "F14" be encoded as "blue jay" or "starling," let "F15" be "hawk" or "dove," and so on. Then, in our code, (15) and (16) are synonymous, and both mean "Henry believes that F15s are F15s."

(15) Henry believes that hawks are hawks.
(16) Henry believes that hawks are doves.

This time it makes *utterly no sense* to suggest that while (15) is true, (16) might be false. Of course, other speakers of English may think we are crazy if we say "Henry believes that hawks are doves," but that is irrelevant to the truth of (16). So are Henry's experiences with the words "hawk" and "dove." For (16) is a sentence *we* are using to describe what Henry believes. And the meaning we have assigned to "hawk" and "dove" gives (16) *exactly* the same meaning as (15). *If* "mortician" has exactly the same meaning as "undertaker," then it is hard to see how the relationship between (9) and (10) could differ in any relevant respect from that between (15) and (16), unless (10) is interpreted metalinguistically.[28]

Stipulations and codes undermine Soames's attempt to explain away the discrepant intuitions in the Mates examples by appeal to a pragmatic principle of fidelity.

> Although the (a) sentences may appear to differ in truth value from the (b) sentences, this appearance is due to pragmatic considerations, most notably, the requirement that the reporter be maximally faithful to the words of the agent unless there is reason to deviate. Since in cases like this there often is no such reason, utterances of these sentences will suggest to the hearer that the reporter has been maximally faithful to the agent's own words. In these particular examples this suggestion takes on added significance because of the triviality of the propositions semantically expressed by the complement clauses. Thus it is natural to regard utterances of (47b) ["A says (believes) that physicians are doctors"] and (48b) ["A says (believes) that Hesperus is Phosphorus"] as incorrect when the suggestions are false. Such utterances are incorrect, but that does not mean that the propositions semantically expressed by these sentences are false. (Soames 1987a: 123)[29]

We will ignore the problem that the speaker may never have expressed the beliefs that we are describing, or may always have expressed them incorrectly. When I describe what Henry believes using my own personal stipulations, technical terms, or codes, I am not being maximally faithful to his words. I may have no reason for using my stipulations other than

28 Note that *Henry believes that "Hawks are doves" is true iff Henry believes that Hawks are doves* would be *false* when interpreted as a sentence in our code, if we assume that Henry does not know the code. For Henry believes that F15s are F15s (making the right side true) but does not believe that the sentence "Hawks are doves" is true (making the left side false). He believes the sentence is false in English, and has no beliefs about our code. Compare Yagisawa 1984: 413–15; contrast Burge 1978: §2.

29 Loar (1987: 182) also suggests that "a belief ascription's being *de dicto* is a matter of its capturing relatively *superficial* facts about the belief – how it would be expressed . . . ," but sees this factor as semantically relevant. See also Noonan 1980–1: §3.

that I feel like it. Yet a sentence like (16) will not seem false (unless it is misinterpreted as a sentence of English) or in any way pragmatically inappropriate. This will be particularly true if I am writing in my private diary.

Soames's pragmatic principle is also refuted by our ability to use a foreign language to describe what an English speaker believes and to do so without any sense of falsity or pragmatic impropriety – even if we could just as well use English. Soames is clear that his pragmatic desideratum is not absolute and may be outweighed by other factors. He cites the use of other languages as such a factor, but he does not explain why, if we are supposed to be maximally faithful to the speaker's own words, the totally optional use of a foreign language would be a legitimate excuse for violating the rule.

> When reporting from the perspective of a third party, we need some excuse for not using the agent's own words (or strict translations of them). (Soames 1987a: 119)

The parenthetical alternative presumably allows for the use of foreign languages. But it also gives up the whole game. A strict translation is just an exact synonym. If "mortician" and "undertaker" are strict synonyms, then any strict translation of one will also be a strict translation of the other.

Note finally that if what Henry says is "undertakers are undertakers," and what I report is "Henry believes that morticians are morticians," I have deviated from what Henry said – for no reason, perhaps – and yet this report will not seem false at all. So even if there were the sort of pragmatic rule that Soames described, its violation would not explain the discrepant intuitions in the Mates objection. I believe a survey of the evidence will show that fidelity to the speaker's own words is relevant to the propriety of a that-clause only when it is relevant to the truth or falsity of the propositional attitude so ascribed. Thus features of the speaker's words such as accent, pitch, tempo, and font are completely irrelevant. The only thing that matters is their content.[30]

When we take clear *non*synonyms, Theorem 13.4 is also confirmed. Consider "square" and "triangle," "number" and "integer," and "square

30 Salmon's "Gricean Strategy" has also been used to defend Millian theories of meaning against apparent substitutivity failures, and will be criticized at greater length in *Nondescriptive Meaning and Reference*.

root of four" and "cube root of eight." When these terms are substituted in belief sentences, the results are clearly nonsynonymous (when the sentences are interpreted as de dicto belief ascriptions).

(17) Henry believes that the square root of four is the square root of four.
(18) Henry believes that the square root of four is the cube root of eight.

If Henry says, "The square root of four is greater than the cube root of eight," we have some evidence that (18) is false. The evidence is not conclusive all by itself, for it may be that Henry said what he did because he mistakenly thinks that "cube root" means "sixth root" rather than "third root." But if he accurately defines his terms, and explains that he reached his conclusion through the use of a calculator, then we would have even stronger evidence that (18) is false. We would have no evidence, by contrast, that (17) is false.

We have not analyzed the notion of similarity of meaning or degree of synonymy in this work, and have only a rough ordering at best anyway. But it seems reasonably clear that the following pairs are listed in order of increasing similarity of meaning: *disease, square; square, triangle; number, integer; undertaker, mortician; premise, premiss.* When these pairs are placed in the form "That is a(n) ____," the result is a sequence of sentence pairs listed in order of increasing similarity of meaning. It seems evident that the more similar in meaning "p" and "q" are in this series, the more similar the proposition that p is to the proposition that q, and the harder it is to imagine people believing one without believing the other. These facts indirectly support Theorem 13.4, which implies that the degree of synonymy between "p" and "q" is at the limit (exact synonymy) when and only when the degree of similarity between the propositions expressed is at the limit (numerical identity).

To sum up, when it is clear that "p" and "q" are synonymous, it is clear that "the proposition that p" and "the proposition that q" are synonymous, along with "S believes that p" and "S believes that q." The more similar "p" and "q" are in meaning, the harder it is to perceive a difference between "S believes that p" and "S believes that q." When it is clear that "p" and "q" are nonsynonymous, it is clear that the expressions containing them are nonsynonymous. When it is unclear whether "p" and "q" are exactly synonymous, it is unclear whether the expressions containing them are synonymous. In that case, it will be unclear whether the proposition that p is identical to the proposition that q, and whether anyone who believes one also believes the other. This is exactly what

we should expect given Theorem 13.4. If, as appears to be the case, the semantic difference between "S believes that p" and "S believes that q" is always more evident intuitively than the semantic difference between "p" and "q," then the conclusion to draw is that *"that" clauses serve as a magnifier for semantic differences.* They help us to discern that very close synonyms are not exact synonyms.

14

The Constituency Thesis

We have asserted on several occasions that thoughts are complex mental events. They have components that occur together in certain relationships when the thoughts occur. The thought that today is Monday and tomorrow is Tuesday has as a component the idea that today is Monday. The thought that the sky is blue has as components the idea of the sky and the idea of being blue. Hence thoughts are composed of ideas, and thinking is ideation. The assumption that thoughts are complex played a central role in developing the expression theory of meaning. The recursion clause of the neo-Gricean analysis depends on it, as does the theory's ability to account for subsentential word meaning. In the next chapter, we will use the fact that ideas (or equivalently, concepts) are thoughts or parts of thoughts in order to define "idea." While widely affirmed,[1] the thesis that thoughts literally have parts has been questioned by many and vehemently denied by some.[2] Since the postulate is foundational to everything else

1 See, e.g., Leibniz 1709: 7.1; Frege 1919; 1923: 55–7; 1979: 143, 150, 253; Geach 1957a: Chapter 14; Cocchiarella 1984: 334–5; Peacocke 1983: 62–4; 1986: 1–2, 15, 63; Jackendoff 1989; Crimmins 1992: Chapter 3; Gillett 1992: 12; Fitch 1993: 471; and all the "language of thought" advocates cited in note 1 to Chapter 20. See also Lyons (1977: 317) on "componential analysis" in semantics, and McCawley (1994) on "generative semantics."

2 Cf. Loar 1981: §9.1; Dennett 1982: 91; Marconi 1990: 86; Ryle 1951; 1958; Geach 1957a: Chapter 14; Evans 1982: 101; Peacocke 1983: Chapter 8; 1986: 63, 114; van Gelder 1990; 1991a; Clark 1991; Bealer 1993a: 30; and, most radically, James 1890: chapter 9, pp. 236, 276–8. There is also a debate among structuralists between those who hold that thoughts are "intrinsically structured" and those who hold that thought is an "undifferentiated continuum" on which language imposes structure (see Lyons 1977: 240–5). But taken literally, this issue concerns the source of thought structure and not its existence, and so is neutral on the issues we are addressing. It is hard to see, though, how a structure could possibly be imposed on a medium that was intrinsically undifferentiated. I would suggest that the real issue here is whether or not thoughts have *discrete* parts – that is, whether the ideas we distinguish in

in this book, and central to cognitive psychology, this chapter will be an extended argument for the constituency thesis. We will observe in passing that while the thesis is plausible for thoughts and the objects of belief, it fails for the act of belief and other propositional attitudes.

§14.1 IDEAS AS THOUGHT-PARTS

The relationship between words and ideas is like that between sentences and thoughts (§13.1). Words express ideas, just as sentences express thoughts. Moreover, words express the same idea if they are synonymous, and meaningless words express no ideas (§13.6). We can even say that the idea expressed by a word is generally contained in the thoughts expressed by sentences containing that word. We can therefore deepen the sentential analogy by adding the following proportion, bearing in mind that ideas and thoughts also differ in many ways from words and sentences.

The Sentential Analogy: *ideas : thoughts :: words : sentences*

It is natural to infer that ideas are parts of thoughts, just as words are parts of sentences.

The term "idea" is more general than "thought." While there is a natural tendency to use "idea" predominantly for *proper* parts of thoughts, it is perfectly proper to refer to the idea that the sky is blue, which is a thought. The idea of the sky and the idea of being blue are not themselves thoughts, however, for they are not suitably complex. Thoughts must have at least two components, related in certain ways. In virtue of the relations among their constituents, thoughts have a structure. We will refer to the basic claim that thoughts have parts as the *constituency thesis.*

14.1 **Postulate:** *Thoughts have parts.*

I use "constituent," "component," and "part" interchangeably. This section will be devoted to clarifying what the constituency thesis says. The remainder of the chapter will seek to prove it.

The simplest thoughts, it seems, are expressible by subject-predicate sentences. The subject expresses one idea, the predicate another. The minimum number of ideas in a thought is probably greater than two. Even a thought as simple as "Tasha is white" contains the idea of the present time in addition to the idea of Tasha and the idea of white. But it

them are in some way "natural" segments. As noted in §14.1, the claim that thoughts have distinct parts (the constituency thesis) does not entail that they have discrete parts.

will suffice for us to note that the lower limit is greater than one. There is no upper limit on the complexity of propositions. A sentence like "One is a number and two is a number . . . and n is a number" is meaningful no matter how large n is, and so expresses a proposition. There is presumably a limit, though, on the complexity of propositions that can actually be thought by any particular subject. But such a limit cannot be established a priori.

Since the occurrence of a complex event entails the occurrence of all of its parts, the constituency thesis together with the definition of ideas as thought-parts (see Definition 15.1) entails what I call the *second law of occurrence*.

14.2 **Theorem:** *A thought occurs to S only if all of the ideas composing that thought occur to S.*

The thought that the sky is blue or green occurs to S only if the thought that the sky is blue occurs to S, and that occurs to S only if the idea of the sky occurs to S. More generally, any complex concept occurs only if its components do. The idea "green ball" occurs to someone only if the ideas "green" and "ball" do. The converse ontological principle fails: the parts of a complex event often occur without that complex occurring. Only if all of its parts occur in the proper relationships will the complex event occur. A sequence containing just the first three notes of a C-major scale is not a C-major scale, for example, nor is a sequence containing all of the notes in random order. Similarly, the idea of the sky can occur to us without the thought that the sky is blue occurring to us. And unless the components of the thought that John loves Mary occur in the right relationship, the thought that Mary loves John, or the idea of John's loving Mary, might occur to us instead.

In the case of ideas and thoughts, however, a limited version of the converse principle of occurrence appears to hold. It is hard to find an idea occurring that is not occurring as part of some thought, unless the idea is a thought itself. For example, someone's saying just the name "Franklin Roosevelt" automatically makes me think of Franklin Delano Roosevelt, but does not automatically make me think any particular thought about Roosevelt. Nevertheless, whenever I think of Roosevelt in this situation, I also seem to have a number of fleeting thoughts about him, such as that he was president, that he was married to Eleanor, and so on. I cannot describe any situation in which it is clear on the basis of either introspection or indirect evidence that I am thinking of Roosevelt without thinking any thought about him. It thus appears to be universally true that *ideas occur*

only as parts of thoughts (proper or improper).[3] Even though we are going to define an idea as a part of a thought, it appears that we could also reverse the order of definition and define a thought as an idea that is complete in the sense that it can occur without being part of another idea.

However, the occurrence of a thought-part does not logically entail the occurrence of a thought containing it, and the evidence that ideas occur only as parts of thoughts is not conclusive. While it is probably true a priori that every idea-*type* is part of some thought-type, the proposition under discussion is the empirical thesis that every idea-*token* is part of some thought-token. Since none of the theses that I wish to establish depends on whether or not ideas can occur independent of thoughts, I will leave this question open.

The constituency thesis says nothing about the *sort* of parts that thoughts have. In particular, the thesis does not say that ideas are *spatial* or even *temporal* parts of thoughts. Since mental events do not appear to be spatially extended, they do not appear to have spatial parts. Of course, it may be that thoughts really are arranged in space, and do not appear to be so only because introspection is limited. The hypothesis that a thought is a two- or three-dimensional neural process cannot be rejected on the basis of introspection alone. There are many other possibilities, though. A thought might have the nonspatial structure of a chord, for example, or of a musical phrase. The chord analogy is especially apt in one respect. For although trains of thought are temporally extended, with one thought occurring after another, the ideas comprising an individual thought appear to occur simultaneously rather than seriatim. Again, this appearance may be due to the limits of introspection. There are further possibilities. For example, in addition to written sentences (of which words are spatial parts) and spoken sentences (of which words are temporal parts), there are "abstract" sentences that can occur either as written or as spoken, and of which consequently words are neither spatial nor temporal parts. There is no reason to assume, however, that thoughts are like even abstract sentences in being *concatenations* of their constituents.[4] In general, a concatenation is a *linear* or *serial* structure, a *sequence* – something that can be represented by an ordered n-tuple of its constituents. Like chemical molecules or musical phrases, thoughts seem to be much too complex to be described as simple linear arrangements of concepts. And there is no

3 Cf. Geach 1957a: 14; Vendler 1972: 41; 1977: 59; Evans 1982: 102; and see §14.3.

4 Compare and contrast van Gelder 1990: 359–61; 1991a: 363–7. While van Gelder's examples are true concatenations, he uses the term "concatenation" to denote any systematic structure.

reason to assume that the parts of a thought are next to each other in any spatio-temporal location. In observing that the relationship between ideas and thoughts is like that between words and sentences, we can claim only an analogy, not an identity. The fact that ideas are contained in thoughts while words are contained in sentences establishes an analogy even if the modes of containment are different.

The constituency thesis presupposes that the parts of thoughts are *distinct*, but not that they are *discrete*.[5] To make this distinction vivid, note that North and South Dakota are distinct geographic parts of the United States, but not discrete parts. Geographically, the lower forty-eight states constitute a continuous geographic region. Similarly, the top half of this X is distinct from the lower half, even though they are not separated. Finally, the surprising fact that the stream of sound produced when a sentence is uttered does not have the discrete segments that native speakers think that it has does not imply that the stream lacks distinguishable segments. Nothing that I say either requires or precludes the assumption that thoughts are combinations of discrete parts. Introspection may reveal that the thought that John loves Mary contains various ideas, and that the thought that Mary loves John is a different thought containing the same ideas. But introspection does not disclose the detailed inner structure of either thought. A complete theory of thought structure will therefore have to be based on indirect evidence from logic, linguistics, psychology, and neurophysiology.

Ryle (1962) assumed that concepts are not *separable* factors in thoughts, since, as we have just noticed, we do not seem to be able to conceive the concept of a planet, say, all by itself, separate from any thought containing it. Let us grant this assumption now for the sake of argument. Ryle concluded that concepts are not occurrences or "incidents" and that there is no such act as conceiving a concept. If this were a valid argument, we could prove by similar reasoning that speech sounds are not occurrences, and that there is no such act as pronouncing the "s" in "bust." But if we never pronounced the "s," the word "bust" would sound just like "but." Note that an inseparable part might be discrete as well as distinct. Even if the forces binding protons and neutrons together in a nucleus were so strong that they could never be separated from each other, there could still be relatively large spaces between them. They would be in separate places even though the distance between them could not be increased.

5 Compare and contrast James 1890: Chapter 9, pp. 237–9, 276–7, 279; Titchener 1909: 26–7; Humphrey 1951: 11; and Pylyshyn 1980: 159.

Peacocke (1986) defined concepts as components of thoughts, but saw a problem.

> The problem concerns the relative priority of thoughts and the constituents from which they are built up. Each content [thought] is conceived of as structured. . . . The problem is that we seem to have no conception of what these constituents, the modes of presentation, are apart from their role as constituents of thoughts. (Peacocke 1986: 63)

Peacocke saw a marked distinction here between thought components and molecular components.

> We *can* conceive of an atom existing independently of its being a constituent of a molecule; and the idea of its existing does not seem to need explanation in terms of its potential role as a molecular constituent. (Peacocke 1986: 63)

This is not a genuine problem. In order to think that something has parts, we do not have to have any substantial conception of what the parts are apart from that complex. Indeed, when atomic theory was first developed by Dalton, he had little conception of what atoms were apart from their role as constituents of solids, liquids, and gases. Adoption of the hypothesis that matter has constituents was essential to discovering the intrinsic nature of atoms. True, Dalton did hypothesize that atoms were material objects, with all of the inherent properties thereof, such as spatial position. Analogously, in thinking of concepts as constituents of complex events, we are conceiving of them as events, with all the inherent properties of events, such as temporal location. We are also conceiving of ideas as mental events that occur to animate beings in various locations. Nothing rules out, a priori, our learning a lot more about the intrinsic nature of ideas. For example, we might discover that thoughts are complex neural activation patterns, and that ideas are constituent activation patterns. Even if concepts never do occur independent of any thought, either by chance or by law, it is not inconceivable – in the way that it is inconceivable for an angle to exist without the lines that form it, or for a thought to occur without its component concepts – that they should. Finally, even if it were inconceivable for concepts to occur independent of thoughts, it is hard to see why that would cast any doubt on their reality. The fact that a triangle cannot exist without its sides casts no doubt on the reality of triangles. And there is nothing to stop a physicist from postulating that even though electrons, protons, and neutrons are the smallest material objects that can exist independently, they nevertheless have constituents. Indeed, I believe that that is what quarks are.

As we observed in §12.3, Peacocke saw a metaphysical problem in the Fregean idea that thoughts (and concepts generally) are abstract objects distinct from token mental states, yet empirically significant. On my view, thoughts are types of mental events, so no such problem arises. Types have causes and effects because their tokens do. The abstract nature of thoughts presents no problem for the thesis that thoughts have parts, for types can have parts as well as tokens. The water molecule is a type of molecule, which contains two atoms of hydrogen and one of oxygen. Sentences provide an even closer analogy. Peacocke thought there was a difference between thoughts and sentences on this score.

> In the case of the pure theory of expression types, linking the concatenation operation on expression types with empirical application is trivial: an instance of the expression type **A**^**B** is any token that consists of an instance of **A** followed by an instance of **B**. Nothing analogous holds in the pure theory of concepts. In the sense in which concepts have instances, the way in which the predicational combination of two concepts is applied in classifying mental states is *not* given by some operation on the objects to which those concepts apply! (Peacocke 1992: 115)[6]

Peacocke is certainly correct in saying that the predicational combination of two concepts is not to be explained in terms of an operation on the objects to which the concepts *apply*. For application (being true of, denoting) is a relation distinct from occurrence (instantiation, tokening). The concept of being far away *applies to* the Sun, not to the Sun concept. That is, it is the Sun that is far away, not the Sun concept. The concept *occurs to* people who are thinking of the Sun. The concept occurs to a subject iff a token of the concept is occurring to the subject, just as a sentence-type occurs on a page iff a token of the type occurs on the page. The concept of being far away and the concept of the Sun are *connected to* each other in the thought that the Sun is far away. But the two concepts do not apply to each other. So we can perfectly well maintain that two concepts are concatenated or otherwise connected iff tokens of the concepts are concatenated.

On my view, the thought that France is west of Germany is a type that occurs to many people at different times. Its occurrences are tokens of the type. The thought-type is a whole composed of the idea of France, the

6 See also Peacocke 1992: 184. Smith and Medin (1981: 7) similarly say that the sentence "Fish are friendly" expresses a proposition in which the concept "friendly" is *predicated of* the concept "fish." A proposition doing that would be true only if the concept of a fish were friendly.

idea of Germany, and the idea of being west of. These ideas are event-types that occur in the thought. Tokens of the thought are also wholes composed of tokens of its component ideas. The part-whole relationship applies to both types and tokens, but has somewhat different properties in the two cases.[7] For example, the rule that wholes are identical iff their parts are identical holds only for tokens. The two sentences "John loves Mary" and "Mary loves John" are two different sentences composed of exactly the same words. Tokens of the two sentences will necessarily contain different tokens of the words, but the sentence-types themselves contain exactly the same three word-types. Consider "isomers." "In chemistry the word is used to designate two or more compounds that contain the same (equal) atoms but differ in the arrangement of these atoms" (Yoder et. al. 1975: 160). The two isomers of CHBrClF (bromochorofluoromethane) contain the same five atoms, but differ depending on the spatial arrangement of the atoms. As a result, the isomers polarize light differently. In general, the identity of a complex type depends not only on the identities of its components, but also on the relationships among its components. Whether we are dealing with types or tokens, different relationships among the components produce different compounds.

Ryle (1951, 1953, 1958) characterized the constituency thesis as a "para-chemical" hypothesis that resulted when Cartesians and Lockeans tried to describe "the functions of bits of theoretical discourse," a conceptual activity, but couched their descriptions as "narratives of introspectible mental processes," an empirical activity. No such error will be found in the evidence provided here. Ryle also pointed to the great variety of ways in which thinking occurs. What is there in common between doing mathematical calculations out loud, thinking up a melody in one's head, and fantasizing? Ryle's evidence does show that visual images and inner-speech words are not literally components of thoughts, since thinking can occur without them. But the great variety of thought processes does not show that thoughts have no components at all – that is, that thoughts are not complex.

Ryle (1958: 192–3) observed that while every schoolgirl can describe her thoughts, no one can say what the having of ideas or thoughts consists in. He thought that this inability was an embarrassment to the

7 Contrast Mates (1986: 62), who erroneously inferred that "the relation between concepts and their components will no longer be that of whole to part," and concluded that talk of containment among concepts was a mere "projection of the containment relations among the corresponding linguistic expressions." An extension of his reasoning would lead to the startling conclusion that the containment relation does not apply to expressions either!

constituency thesis. However, the fact that children are equally adept at describing water and wood and glass even though they are completely unable to say what these things consist of does not even begin to show that household materials do not have constituents. Ryle's argument seems to assume that any facts about our thoughts must be known directly, by simple introspection. This assumption does not follow from the fact that thoughts are introspectible. Water is visible, after all, even though its constituents are not. The constituents of benzene were known for a long time before its structure was. Furthermore, the supposed omniscience of introspection is undermined by independent considerations. The time at which a thought occurs, for example, is not something that is known directly by introspection. The evidence that I will provide for the constituency hypothesis is indirect, but that does not make it inconclusive.

My final preliminary observation is that in taking ideas or concepts to be parts of thoughts, we are not presupposing that the set of concepts is either *fixed* or *innate*.

> The principal difficulty with the view that cognitive activity consists essentially in the rule-governed manipulation of a combinatorial lexicon of internal representations is that it is inadequate to explain the most impressive aspect of the most important kind of cognitive activity: large-scale learning. The most impressive aspect of the learning process, whether seen in the evolution of scientific theory over centuries or in the human child's conceptual development over its first few years, is that the lexicons of internal representations used by scientists and infants display the characteristic tendency to evolve and change, often radically and sometimes totally; and so do the apparent rules for manipulating them. . . . [L]arge scale learning appears to be identical with conceptual change. . . . It will be argued that such change as I have cited is change at a relatively superficial level: that it can be accounted for in terms of the articulation in context of a deeper set of fixed rules operating on a fixed lexicon of more primitive representations, from which the more complex and fluid representations are formed by suitable concatenation and re-concatenation. [But] conceptual change is not plausibly explained by the recombination of a fixed set of primitive or simple concepts; not (just) because we have failed to identify the lexicon of such an innate 'language of thought,' but for the prior reason that our manifest concepts appear not to 'decompose' in the manner required by such a view. (P. M. Churchland 1980: 133–4)[8]

The simple claim that thoughts have components does not entail that a person's thoughts do not change over time, nor that the set of atomic or

8 For an example of Churchland's target, see Jackendoff 1989: 98.

primitive concepts does not change over time. It certainly does not entail (or deny) that any concepts are innate. *Pace* Fodor, it is patent that both the child and the scientist acquire a multitude of new concepts, often through abstraction and other experiential processes (see Chapter 17). A child normally does not acquire the concept of a cosine until age twelve or so, and scientists did not form the concept of a quark until the twentieth century. Finally, nothing that we have said entails that ideation is the only type of cognitive activity. Hence the constituency thesis is fully compatible with the existence of "paradigm shifts" in our conception of the world, which involve systematic changes in belief rather than in mere thought (see §19.4).

§14.2 THE CONSTITUENCY THESIS FOR THOUGHTS

Having clarified what the constituency thesis claims, we turn to establishing its truth. Consider the following thoughts.

TvW The thought that today is Tuesday or Wednesday.[9]
TvM The thought that today is Tuesday or Monday.
T The thought that today is Tuesday.

It is undeniable, I believe, that the thought that today is Tuesday or Wednesday cannot occur to a subject unless the thought that today is Tuesday occurs to him. It is also evident that these two thoughts are similar to each other in a certain respect and to the thought that today is Tuesday or Monday; for example, all three thoughts are about Tuesday. But such external relationships as co-occurrence and similarity do not logically entail any internal structural relationships, and it may not be evident that T is literally a component of TvW and TvM. What reasons are there to believe that thoughts have a constituent structure? Unless we can provide reasons, it will inevitably be thought that we are just uncritically assuming that thoughts have a constituent structure, or that we are using mereological terms like "part" and "structure" metaphorically. But since the constituency thesis seems to be self-evident, at least to me, it is hard at first to find supporting arguments that do not beg the question. Some arguments are useless for our purposes, such as the Leibnizian inference that ideas are complex because the objects of our thought are complex, which

9 We will be using symbols like "TvW" as *names of thoughts* rather than as *sentences expressing thoughts*. Thus "TvW" is an abbreviation for "the thought that today is Tuesday or Wednesday" rather than for "today is Tuesday or Wednesday."

is valid only on the problematic Cartesian definition of ideas (see §19.1).[10] Fortunately, I believe that supporting arguments can be constructed, in abundance.

We should note first that to show that *thoughts* are structured, it does not suffice to show that *thought* is structured.[11] "Thought is structured" means that the totality of what we think or believe has a structure. This is true because the set of propositions that we believe stand in logical and other relations to each other, and would remain true even if the individual propositions themselves had no internal parts. The fact that we form beliefs according to certain rules of inference entails that our thought has a structure, but does not entail that any particular thought has a structure. For example, thought would be structured in an important way if the subject believed the logical consequences of anything he believes, so that if he believed T he would also believe TvW. It does not follow that either TvW or T has an internal structure.

For similar reasons, the fact that thoughts have complex acceptance or truth conditions does not entail that they have internal structure. The fact that the truth of T is a sufficient condition of the truth of TvW and TvM, for example, does not show that T is a part of TvW or TvM. This seems to have been denied by Peacocke.

> On the present acceptance-condition theory, for a thought to *have* a certain constituent just *is* for it to stand in certain relations to other thoughts and mental states: for thoughts with canonical links, the relations are those given in the canonical grounds and commitments for thoughts of the type containing that constituent. . . . The talk of composition is just a vivid way of encoding more or less complex intrinsic features of the patterns of canonical grounds and commitments which ultimately individuate the thought. (Peacocke 1986: 63)

As Peacocke conceived them in *Thoughts*, canonical grounds and commitments are acceptance conditions (conditions under which a proposition ought to be believed) that constitute truth conditions (conditions under which a proposition is true). The fact that a thought has such truth conditions is perfectly compatible with the claim that the thought has no

10 Leibniz 1676: 283;1709: §2.2.1. Reversing the Leibnizian error, James (1890: Chapter 9, pp. 276–8) concluded that thoughts are undivided from the fact that a thought does not contain one idea for each part of the object of thought. However, even James speaks of parts of thoughts when trying to describe and explain mental phenomena elsewhere (e.g., 1890: Chapter 14, pp. 579–81).

11 Contrast Gillett 1992: 15.

components at all.[12] Suppose, for example, that P is true (and acceptable) iff R and S are true (and acceptable). It does not follow that P is composed of R or S. P could, for all we have said, be a statement constant of sentential logic, one introduced as an abbreviation for the conjunction R&S. P would then have the truth conditions given without having any internal structure. There is no reason to think that the differences between thoughts and symbols make the move from complex truth conditions to internal structure any more valid for thoughts.

A more defensible view is suggested by the final sentence just quoted, and that is the view that talk about the components of a thought is just a manner of speaking about its features, especially its relational features. On this view, thoughts have features but do *not* have constituents. We may *say* that they have constituents, but this is just a *façon de parler* and is not literally true.[13] Obviously, this word game is useless for our purposes. We need to show, on the contrary, that individual thoughts *literally* have constituents that form a structure.

It is possible to hold that one thing is closely related to another without holding that one is a part of the other. Indeed, there are three possible alternatives to the constituency thesis for thoughts: the *feature*, *genus*, and *correlation theses*. The *connectionist thesis* is a special case of the correlation thesis. We will discuss each of these alternatives in turn, and show why the constituency hypothesis is superior.

The Feature and Genus Theses

Apples are intimately related to their size, shape, weight, and color, and their relationships to other objects. These are an apple's features – its properties or attributes, not its parts. If this were the right model, then the apparent components of a thought would be features of the thought rather than genuine components. They would have to be essential properties, we might add, since a disjunctive thought cannot occur to a subject without its disjuncts. The feature hypothesis for thoughts can be ruled out on metaphysical grounds, however. Events and their features belong to different ontological categories. This means that the sorts of external relationships noted here could not obtain if the feature theory were true.

12 Peacocke (1983: 209–10) himself showed that unstructured thoughts could be caused by sets of "recognitional abilities."

13 Cf. Frege 1923: 55; Evans 1982: 108, fn. 31; van Gelder 1990; 1991a; Bealer 1993a: 22. Peacocke maintains his view in *A Study of Concepts* (1992: 105–24).

The features of an event cannot be very similar to the event. The features of a thought or idea occurring to S are not themselves thoughts or ideas occurring to S, and more generally, the features of an event occurring to S are not themselves events occurring to S. *Being closely related to the thought that today is Tuesday* is a feature of the thought that today is Tuesday or Wednesday, as are *being distinct from, being similar to, being implied by,* and *having more chance of being true than* the thought that today is Tuesday. But *the thought that today is Tuesday* is not a property of any thought. The latter may occur to or be thought by a person, but none of the attributes can. A thought may possess many properties, but there is no sense in which one thought or idea may possess another.

The metaphysical problem of the feature theory can be avoided by holding that the thought that today is Tuesday is a *genus* of which the thought that today is or is not Tuesday is a *species*. The apparent components of a thought, on this view, are generic events, and the given thought is a specific event of that type. Any occurrence of a generic event is an occurrence of one of its species, and vice versa. However, the genus thesis is a category mistake for different reasons. The thought that today is or is not Tuesday is not itself an instance of the thought that today is Tuesday in the way that leukemia is an instance of cancer. How could a tautologous disjunction be an instance of a contingent subject-predicate proposition? Since the tautological disjunction is related in the same way to both the thought that today is Tuesday and the thought that today is not Tuesday, the genus thesis implies further that one consistent thought could be an instance of different and contradictory thoughts, which is absurd. Finally, the thought that today is Tuesday could not be a species of the idea of Tuesday, because the latter is not a type of thought. There is of course a sense in which the thought that today is a rainy Tuesday is more specific than the thought that today is Tuesday. But that describes the contents of the thoughts. And in the same sense, the thought that today either is or is not Tuesday is not more specific than the thought that today is Tuesday.

The Correlation Thesis

According to the correlation thesis, the apparent components of a thought are ideas that are separate yet always occur when it does. The first consideration favoring the constituency over the correlation thesis is *simplicity*. Since the correlation hypothesis entails that the thought that today is Tuesday or Wednesday is something that exists over and above the complex of distinct thoughts and ideas that occur with it, it is ipso facto a

more complex hypothesis. Both the constituency and the correlation hypotheses grant the existence of a complex of distinct but co-occurrent thoughts and ideas: the mereological sum containing among other things the ideas of today, of Tuesday, and of Wednesday, the thought that today is Tuesday, and the thought that today is Wednesday. But according to the correlation hypothesis, the thought that today is Tuesday or Wednesday is an *additional* entity. (*Analogy*: The theory that a water molecule is separate and distinct from the two atoms of hydrogen and one of oxygen that are always present with it postulates more complexity than does the theory that the water molecule consists of the three atoms.)

The second and most important consideration favoring the constituency hypothesis is *explanatory power*. It explains many facts that are left unexplained by the correlation hypothesis, including: introspective integration, similarity relations, co-occurrence relations, systematic relations, systematicity and productivity, referential dependencies, the compositionality of linguistic meaning and reference, and the phrase-structure syntax of natural language sentences. We will describe each of these phenomena, and show that only the constituency thesis explains them.

Introspective Integration. When the thought that today is Tuesday or Wednesday occurs to us, the thought that today is Tuesday occurs to us, as does the thought that today is Wednesday and the concept of disjunction. Contrast this with a case in which I am listening to a piece of music while wondering what day it is. Suppose I am thinking "That is an oboe" while thinking "It is Tuesday or Wednesday." Introspectively, the relationships among the first set of thoughts seem to be very different from the relationships among the second set. In the first case, I seem to be aware of one unified thought process. The thought that today is Tuesday or Wednesday does not seem to be something taking place over and above the occurrence of the thought that today is Tuesday, of the thought that today is Wednesday, and of the concept of disjunction; and the latter three occurrences do not appear to be separate and coincidental. In the second case, there seem to be two separate thought processes going on simultaneously but coincidentally. Knowing what it is like to think "Today is Tuesday or Wednesday" includes knowing what it is like to think "Today is Tuesday," but not knowing what it is like to think "That is an oboe."

The constituency hypothesis, according to which thoughts really are composed of their apparent components, provides a simple explanation of the introspective integration we find. The thought that today is Tuesday or Wednesday does not appear to be something taking place over and above

the thought that today is Tuesday, the thought that today is Wednesday, and the concept of disjunction because the disjunctive thought is a whole composed of the other ideas. The occurrence of the latter ideas in the proper relation just is what it is for the disjunctive thought to occur to us. The correlation hypothesis provides no explanation for introspective integration. Moreover, it predicts integration where it does not occur. Even though the thought "The flag is at half-mast" is correlated with the thought "Someone died," the two thoughts do not appear to form one unified process when they co-occur. The same goes for the idea of salt and the idea of pepper.

Similarity Relations. The constituency hypothesis also provides a simple explanation of the similarity relations noted earlier. Why are TvW and TvM similar, and why are both about Tuesday? The fact that TvW and TvM both have the thought that today is Tuesday as a component provides a simple explanation. The sentences "Today is Tuesday or Wednesday" and "Today is Tuesday or Monday" are similar for similar reasons. The thesis that occurrences of TvW and TvM are correlated with occurrences of T does not imply that these thoughts are similar in any significant way or that they are about the same things.

Co-occurrence Relations. The constituency thesis also explains the co-occurrence relations we have just noted. Why is it impossible for the thought that today is Tuesday or Wednesday to occur to someone unless the thought that today is Tuesday occurs to him? And why is it impossible for those thoughts to occur to someone unless the idea of Tuesday occurs to him? The natural explanation is just that the idea of Tuesday is part of the thought that today is Tuesday, which is part of the thought that today is Tuesday or Wednesday. The second law of occurrence (Postulate 14.2) provides the covering law for this explanation. On the correlation thesis, these thoughts and ideas are connected for some other reason, or by brute fact. The correlation thesis cannot even explain why the thought that today is Tuesday or Wednesday *does not* occur without that whole complex occurring, let alone why the former *cannot* occur without the latter. The fact that it seems completely silly to try to explain why the former does not or cannot occur without the latter is a sign that there is no good reason to adopt the correlation hypothesis. The correlation hypothesis is on a par with the suggestion that a pianist's playing a C, an E, and a G are merely correlated with his playing the C-E-G chord. (If the feature theory were coherent, it would have similar deficiencies. For we would

have no explanation of why the apparent components of a thought are features of the thought, or of why they are essential features.)

Systematic Relations among Occurrences. More significantly still, the constituency thesis explains why occurrences of the apparent components of a thought are related in certain ways when the thought occurs, and not otherwise. The thought that today is Tuesday and the thought that today is Wednesday are related in one way when the thought that today is Tuesday *or* Wednesday occurs to us, and a different way when the thought that today is Tuesday *and* Wednesday occurs to us. Let us consider the following three thoughts in more detail.

Ljm The thought that John loves Mary.[14]
Lmj The thought that Mary loves John.
Hjm The thought that John hates Mary.

When the thought that John loves Mary occurs, the ideas of John, Mary, and loving occur. But these three ideas do not just co-occur; their occurrences are related in certain ways. The same three ideas also co-occur when the thought that Mary loves John occurs, but their occurrences are related differently in that case. Indeed, the ideas of John and of Mary have switched roles in relation to each other and to the idea of loving. When the thought that John hates Mary occurs, the occurrence of the idea of hating is related to the occurrences of the ideas of John and of Mary in the same way that occurrences of the idea of loving were related to the other ideas when the previous thoughts occurred. The idea of hating has taken the place of the idea of loving. In other words, the apparent components of thoughts satisfy the ideational analogues of the *movement* and *substitution tests* used to show that sentences have phrase structure (O'Grady et. al. 1993: 165). In the case of sentence structure, the tests are satisfied by change of place, and replacement, in a linear spatial or temporal sequence. In other cases, such as musical, chemical, and mathematical structures, the change of place and replacement occur in virtue of different relationships that nevertheless bear a formal similarity.

These facts are easily explained on the constituency hypothesis: the ideas of John, Mary, and loving occur in certain relationships when thought Ljm occurs, because Ljm *is* a compound consisting of those ideas related in certain ways. The same facts remain to be explained on the

14 The symbol "Ljm" is to be interpreted here as a name, not a sentence: it abbreviates "the thought that John loves Mary" rather than "John loves Mary."

correlation thesis. If thought Ljm were a distinct entity, not composed of the ideas of John, Mary, and loving, then we would have to explain why these ideas occur in certain relationships when Ljm occurs but not otherwise. Why should the thought have such an effect on the ideas if they are separate from it? Consider two analogies. The hypothesis that the first chord of Beethoven's *Moonlight* Sonata is correlated with the occurrence of three notes, two C-sharps's and a G-sharp, would not explain why the two C-sharps's are always one octave apart, or why the G-sharp is always one fifth above the higher C-sharp. The hypothesis that the first chord of the *Moonlight* Sonata consists of those three notes in that relation does. Similarly, the fact that the number of hydrogen atoms and oxygen atoms present whenever water is present stand in the ratio of 2:1 is a consequence of the hypothesis that each water molecule is composed of two hydrogen atoms and one oxygen atom. The hypothesis that water is correlated with hydrogen and oxygen does not explain the observed numerical ratios.

Systematicity and Productivity. In a related argument, Fodor often cites the systematicity of thought as the best argument for constituency.[15] "Systematicity" refers to the fact that thinking one thought implies an ability to think other thoughts related to it in certain ways, ways that resemble the systematic syntactic relations among sentences. In particular, thinking the thought expressed by one sentence (or set of sentences) implies an ability to think the thoughts expressed by grammatical permutations of the sentence(s). For example, anyone who thinks Ljm is able to think the three other specific combinatorial possibilities Lmj, Ljj, and Lmm. The fact that psychological abilities are predicted by the combinatorial possibilities is what provides such strong evidence for the constituency thesis. My argument from systematic relations is clearly similar to Fodor's argument from systematicity, but there are differences. I used constituency to explain relations among occurrences; Fodor used it to explain systematic relations among our abilities. I looked at the fact that people often *do* think both Ljm and Lmj. Fodor looked at the fact that anyone who

15 E.g., Fodor 1987: 147–53; 1998a: 97–100. See also Evans 1982: 103–4 (the "generality constraint"); Fodor & Pylyshyn 1988: §3.2; Sterelny 1990: 24–5; Davies 1991: 239–44; Schiffer 1994; Rey 1995: 213. Cummins (1996) believes that Fodor's systematicity argument begs the question. But that is because he formulates the principle of systematicity by *mentioning thought forms* rather than using or mentioning *sentences* of different forms in referring to different thoughts. A. Clark (1991) dismisses the systematicity argument on the grounds that the principle of systematicity is a conceptual truth, and therefore not something capable of empirical explanation. Besides being dubious, his premise is irrelevant. The constituency thesis is supported by systematicity whether it is a conceptual or empirical truth.

thinks Ljm *can* think Lmj. Hence I focused on "performance," Fodor on "competence."

Other propositional attitudes are systematic, too, but none appear to be as systematic as thought. While someone who thinks "2 > 1" is clearly able to think "1 > 2," it is doubtful that anyone who believes the former is able to believe the latter, unless "able" is taken to express a very weak type of possibility that exists unconditionally. It is logically possible for anyone to believe that 1 > 2, for example. But this possibility is extremely remote, and exists whether or not the person believes that 2 > 1.

Although exceptions to the systematicity of thought are hard to find, it is unclear just how general even it is. In particular, Sterelny (1990: 183) and Dennett (1991: 27) worry that systematicity fails for nonlinguistic animals. There must surely be some limits even in humans, for the systematicity of thought coupled with the fact that people have certain iterative concepts implies *productivity*, the ability to think novel thoughts of unlimited complexity.[16] For example, if S thinks "If Mary loves John then John loves Mary" (A → B), then systematicity implies that S can think all of the thoughts in the following series: "If John loves Mary then: if Mary loves John then John loves Mary" (B → (A → B)); "If Mary loves John then: if John loves Mary then: if Mary loves John then John loves Mary" (A → (B → (A → B))), and so on. Since every human being has finite mental capacities, there must be some thought in this series that is too complex for S to think. A Chomskyan might view S's general inability to think thoughts beyond a certain level of complexity as a "performance limitation," in the absence of which S would be able to think the specific thoughts indicated. But there is no need for us to rescue unlimited systematicity by divorcing "competence" from "ability" in this problematic way. The argument for constituency does not depend on how systematic our abilities are. Indeed, the evidence would be decisive even if thinking were a much more isolated or "punctate" ability than it appears to be. As we have seen, the argument for constituency does not depend on any facts about competence as opposed to performance. The fact that we have actually thought thoughts expressed by long and complex sentences generated by repeated iterations of recursive operations suffices.

Fodor and Pylyshyn assert that "systematicity follows from the postulation of constituent structure" (1988: 121). This appears to be an

16 Several authors offer productivity as evidence of constituency. See, e.g., Fodor 1987: 147–53; Fodor & Pylyshyn 1988: §3.1; Jackendoff 1989: 70–1; Maloney 1989: 8; Sterelny 1990: 24–5; Lycan 1993; Schiffer 1994.

overstatement. Molecules have a constituent structure, but the fact that H_2O exists does not imply that O_2H is a possible molecule (unless we are just talking about logical possibilities). Moreover, we have observed that the systematicity of thought must have limits, and that other propositional attitudes are even less systematic than thought. Fortunately, we do not need to claim that constituency implies systematicity in order to infer constituency from systematicity. It suffices to observe that the facts of systematicity would be a mystery unless thoughts did have constituent structure. Without assuming constituency, we would be hard-pressed to understand why thinking the thought that John loves Mary should be any more closely connected with the ability to think the distinct thought that Mary loves John than it is with the ability to think the distinct thought that the square root of 10 factorial is an imaginary number. With constituency, we can sketch an explanation. If a thought occurs to S, then its components do. This implies that its components are able to occur to the subject. It is the nature of a relational concept that it can occur with individual concepts in either its subject or object position. So if the thought that John loves Mary occurs to S, then the concept of loving should also be able to occur to S with the concept of John in the object position and the concept of Mary in the subject position. Since that is what it is for the thought that Mary loves John to occur, S should be able to think that thought too.

Referential and Representational Dependencies. There are many referential dependencies among thoughts and ideas, where by "reference" I mean extension.[17] The extension of thought Ljm, for example, depends on the extension of the ideas of John, Mary, and loving. The thought that John loves Mary is true iff the ordered pair consisting of the extension of the idea of John and the extension of the idea of Mary is in the extension of the idea of loving. If Ljm were separate and distinct from those three ideas, then we would have to wonder why such a dependency should exist. The existence of such a dependency is not at all surprising on the constituency thesis, given that the properties of a whole are in general dependent on the properties of its parts. We noted earlier that the constituency thesis explains certain similarities in representational content. It is surely not a coincidence that the thought that John loves Mary is about the individuals that the concepts of John and Mary represent. This similarity is to be expected on the assumption that the thought contains the concepts, but

17 See Fodor & Pylyshyn 1988: §3.3 and Crane 1990. Contrast Schiffer 1994.

would be a mystery if the two ideas were separate and distinct. The evidence suggests, in other words, that thought has a *compositional semantics*: that the content and reference of a thought is determined by the contents and referents of its components. This could be true, of course, only if thoughts had components.

The Compositionality of Linguistic Meaning and Reference. The correlation thesis leaves unexplained many facts about how the meaning of a complex sentence is related to the meanings of its components, which facts are readily explained on the constituency thesis. On both the constituency thesis and the correlation thesis, it is possible to provide a semantics according to which the meaning of a compound sentence is a *function* of the meanings of its putative components. The rule that when "S_1" expresses P_1 and "S_2" expresses P_2, "S_1 or S_2" expresses P_3 tells us that the meaning of "S_1 or S_2" is a function of the meanings of "S_1" and "S_2," even if P_3 is not composed of P_1 and P_2. But that rule provides no account or explanation of the fact that the meaning of "S_1 or S_2" is related in an intimate way to the meanings of "S_1" and "S_2." For example, we have no account of the fact that "S_1 or S_2" is about the same things that "S_1" and "S_2" are about, that "S_1 or S_2" is true if and only if "S_1" or "S_2" is true, and so on. On the correlation hypothesis, we could of course postulate that P_3 is about the things that P_1 and P_2 are about, and so on. But we would have no explanation of the truth of such postulates. They would describe mere coincidences. On the constituency thesis, by contrast, we can say that P_3 is about the same things that P_1 and P_2 are about because P_3 *is* the disjunction of P_1 and P_2 (that is, because P_3 is a complex entity consisting of P_1, P_2, and the concept of disjunction in certain relationships). In short, only the constituency hypothesis makes a true compositional semantics possible.

Since a compositional semantics is often defined to mean one in which the meaning of an expression is a function of the meanings of its constituents,[18] let me provide a concrete illustration that will show how weak this formulation is. Imagine a language L with a limited number of color terms C: some mean *red*, some mean *orange*, some mean *yellow*,

18 See, for example, Barwise & Parry 1983: 31; Partee 1984: 281; Fodor 1987: 150; Schiffer 1991: 181, 190; Goschke & Koppelberg 1991: 138; Block 1993: 6; Lycan 1995: 589; Kamp & Partee 1995: 135. Block's (1993: 6) notion of "hyper-compositionality" provides another illustration of how weak the functional definition is (see §8.4). Zadrozny (1994) presents a mathematical proof that *every* semantics can be transformed into one satisfying the functional definition.

Table 14.1. *A noncompositional meaning function*

	red	*orange*	*yellow*	*green*	*blue*
bird	*orange cat*	*yellow cat*	*green cat*	*blue cat*	*red cat*
cat	*orange dog*	*yellow dog*	*green dog*	*blue dog*	*red dog*
dog	*orange fish*	*yellow fish*	*green fish*	*blue fish*	*red fish*
fish	*orange bird*	*yellow bird*	*green bird*	*blue bird*	*red bird*

some mean *green*, and some mean *blue*. The language has a similarly limited number of animal terms A, meaning *bird*, *cat*, *dog*, and *fish*. The language has lots of synonyms, but little expressive power. There might be ten terms meaning *red*, but none meaning *purple*. Now imagine that these color and animal terms combine as follows. "CA" means *orange cat* if "C" means *red* and "A" means *bird*; "CA" means *yellow dog* if "C" means *orange* and "A" means *cat*; "CA" means *green fish* if "C" means *yellow* and "A" means *dog*, and so on. The complete rule is indicated in Table 14.1. Thus in L, "Tasha ets glork noke" might mean that Tasha is a red cat, even though "Tasha ets glork" means that Tashsa is blue and "Tasha ets noke" means that Tasha is a bird. In this language, the meaning of an expression of the form "CA" is a *function* of the meanings of its components. For given any assignment of meanings to its components, there is a unique assignment of meaning to the compound. But this language does not have a true compositional semantics. For the meaning of "CA" is in no way *composed* of and does not include the meanings of its constituents.

True compositionality appears to be a fundamental feature of all natural languages, a "linguistic universal."[19] A noncompositional language is so alien to our conception of language that it is hard for us to think one up, and the result seems very bizarre when we do. The demands that such a language would make on memory would surely make it incredibly difficult to master. Within the framework of an ideational theory of meaning, a compositional semantics would be impossible unless some ideas were composed of others. The way in which this feature reduces the load on memory is easy to understand: once the individual terms "brown" and "cow" are understood, the components of the idea expressed by "brown cow" are already occurrent, and do not need to be recalled or activated.

The Phrase-Structure Syntax of Natural Language Sentences. The unit of a natural language used to express thoughts is the sentence. In English, and

19 Objections to compositionality are addressed in §10.6.

in every other known language, sentences have a phrase structure. Sentences are not random sequences of words. The words are organized into a hierarchical structure of phrases. The correlation thesis provides no explanation for the basic fact that we use phrase-structured symbols to express our thoughts. If the thought that John loves Mary were really separate and distinct from, and yet correlated with, the ideas of John, Mary, and loving, then why would we not use something like an *unordered* quadruple AJML to express the thought that John loves Mary and associated ideas, where A expresses the thought that John loves Mary, J the idea of John, M the idea of Mary, and L the idea of loving? To say that the quadruple is unordered is to say that AJML, LJAM, JMAL, and so forth would all be synonymous and interchangeable. On the other hand, if the thought that John loves Mary were literally composed of the ideas of John, Mary, and loving, then it would be more efficient and informative to express the thought using a symbol whose structure reflects the structure of the thought. The demands on memory would be much diminished.

One answer that the correlation theorist might give as to why we do not use unordered symbol sequences to express our thoughts is that such a language would be either unlearnable or expressively inadequate. The number of thoughts we can think is infinite, which would make it impossible to learn a separate symbol for each. Just as problematic is the *creative* aspect of language use stressed by Chomsky. Since we rarely use the same sentence twice, there is generally no opportunity to learn the associations between the sentences we use and the thoughts they express. A language can be learned, the argument goes, only if there is a recursive scheme for generating all sentences from a finite lexicon and grammar. However, the mere fact that the sentences of a language can be generated from a finite base does not suffice for anyone to understand any of its sentences. The sentences of such a language will have a learnable meaning, by the same reasoning, only if there is a parallel recursive scheme for assigning thoughts to the sentences on the basis of an assignment of ideas to the words in the lexicon. And that would be possible only if each thought were identical to a unique combination of basic ideas. A mere correlation between the thoughts and the idea combinations would not suffice. The correlations could not be learned, since the list of thoughts is infinite, and new sentences will present new idea combinations.[20]

20 This argument, of course, derives from Davidson (1967), although his "semantic values" were truth and reference rather than thoughts and ideas. See also Lycan 1984: 19 and Fodor 1987: 150.

The connectionist thesis is the special case of the correlation thesis on which the apparent components of a thought are correlated because one *activates* the other.[21] Connectionism postulates a network of elementary units or nodes, each of which has excitatory and inhibitory connections to other units. The activation of one unit has a probability of activating other units that is determined by the connections among the nodes. The probabilities can change over time as connections are strengthened or weakened. On the connectionist thesis, the network of thoughts contains at least three nodes:

TvW The thought that today is Tuesday or Wednesday.
T The thought that today is Tuesday.
W The thought that today is Wednesday.

The latter two thoughts are not parts of the former, on this view; they are connected to the former.

Some use the term "connectionism" to denote a framework for explaining behavior without postulating thoughts or other propositional attitudes (e.g., A. Clark 1991: 216). We are using it here to denote a thesis about relations among thoughts. Smolensky (1988: 33–4) appears to propose a connectionist theory that would identify TvW, T, and W with patterns of node activation in which the pattern identified with TvW includes the patterns identified with T and W.[22] But while focusing on activation relations among nodes generally, such a theory would actually embrace the constituency thesis concerning the relations among the three thoughts, not what I am calling the connectionist thesis. Taking a different approach, van Gelder (1990, 1991a) cites Smolensky's work as demonstrating that it is possible for representations like TvW to be "functionally composed" of T and W, even though the latter are not literally contained in the former. Van Gelder uses Gödel numbering to illustrate his notion of functional compositionality. If the Gödel numbers of *A*, *B*, and & are 1, 2, and 3, then the Gödel number of *A* & *B* is 1350.[23] Van Gelder would say that 1350 is "functionally composed" of 1, 2, and 3 because

21 See Quillian 1968; Collins & Loftus 1975; Anderson 1983; Rumelhart & McClelland 1986; van Gelder 1990, 1991a; Horgan & Tienson 1991; Stillings et al. 1995: 28; Macdonald & Macdonald 1995. In addition to the critiques found in Macdonald & Macdonald and Horgan & Tienson, see Johnson-Laird, Herrmann, & Chaffin 1984.

22 See also O'Leary-Hawthorne 1989; A. Clark 1991: 210; Dyer 1991: 404–5.

23 See van Heijenoort 1967: 351.

there is a formal rule that would take 1350 as input and produce 1, 2, and 3 as output, and vice versa. But whereas *A* & *B* is literally composed of *A*, *B*, and &, 1350 is not the mereological sum of 1, 2, and 3 (whether we are talking about numbers or numerals). So the statement that TvW is "functionally composed" of T and W would be an instance of the connectionist thesis, not of the constituency thesis.

Since the connectionist thesis is a version of the correlation thesis, it has most of the drawbacks of the more general thesis. Connectionism is more complex in its implication that an additional entity occurs over and above the complex of apparent components. It holds that TvW is not the sum of T and W (plus disjunction), but something occurring separate from them. Indeed, it holds that the complex consisting of TvW, T, and W is distinct from TvW. The connectionist thesis fails to explain the introspective integration of the apparent components, the similarity of the thought to its apparent components, their apparent satisfaction of the movement and substitution tests symptomatic of syntactical relationships, the systematic relations among thoughts and cogitative abilities, referential dependencies, the compositionality of linguistic meaning, and the fact that linguistic expressions have phrase structures.

The connectionist thesis does provide some explanation for the fact that a thought co-occurs with its apparent components. The co-occurrence is not a mere correlation, but a result of the excitatory and inhibitory connections among the thoughts. But activation is in general a stochastic process: one neural structure, for example, has a greater or lesser tendency to activate another. Only in the most unusual conditions, if any, would one structure invariably activate another. So on the connectionist theory too it needs to be explained why the thought that today is Tuesday or Wednesday never occurs without the thought that today is Tuesday. Connectionism is not capable of explaining why the former *cannot* occur without the latter. It leaves open the possibility that the excitatory and inhibitory connections among the three thoughts just mentioned could change in such a way that they no longer occurred together, which is absurd. Furthermore, activation is a process that takes time. So if the thought that today is Tuesday occurs with the thought that today is Tuesday or Wednesday because the latter activates the former, then there would have to be a finite, if small, time period during which the thought that today is Tuesday or Wednesday is occurring but the thought that today is Tuesday has not yet occurred. But this makes no sense. The connectionist thesis commits the error of classic associationism, which tried to make the relation between the ideas

of salt and pepper the model for all relations among ideas. But the relationship between the thought that today is Tuesday or Wednesday and the thought that today is Tuesday is radically different from the relationship between the thought that today is Sunday and the thought that today is the Sabbath, or that today is Super Bowl day. Associative relations will be the focus of Chapter 18.

The argument providing the basic motivation for connectionism as an alternative to the constituency thesis is a non sequitur.

> In short, the Classical view construes thought as language-like in involving the manipulation of symbols, or representations, that have both a syntax and a semantics. However, in the 1980s another, alternative view of cognition, based on a model known as connectionist, or parallel distributed processing, emerged. A motivating factor in the development of this new conception was to construct a model of cognition that more closely resembles the structure of the human brain, and so is more biologically plausible than classical models. What prompted the conception was the recognition that the brain is a neural network. (Macdonald & Macdonald 1995: xiv)[24]

The brain is indeed a neural network. But it is also a mereological structure. In addition to the synaptic connections between neurons, there are part-whole relationships between neural structures and their components. Complex patterns of neural firing can be analyzed into parts, and are ultimately composed of individual neural firings. Similarly, some cognitive relations are connectionist. The association of ideas (Chapter 18) is a clear example; the support of beliefs and desires by reasons is probably another.[25] Others, as we have argued in this section, are mereological. Cognitive theory need not be a Procrustean bed. We must avoid the false dichotomies that dominate the field.

24 See also Churchland & Churchland 1983; McClelland, Rumelhart, & Hinton 1986: 10; Dennett 1986: 66–7; Bechtel 1988: 30, 39–40, 50–52; van Gelder 1991b: 56; Ramsey, Stich, & Garon 1991; Davies 1991: 250–1, 254–5. Contrast Bechtel 1988: 30–6; 46–50; 53–5; Lycan 1990: 202, 278; Sterelny 1990: §§8.2–8.3; Boden 1991; Garson 1991. Contrast especially Fodor & Pylyshyn (1988: 143–6) and Tienson (1991: 6–7), who note how absurd it would be to argue that a digital computer cannot be processing symbols on the grounds that the computer is an electronic system; and Cummins (1991), who designs a network that performs simplification and other logical operations.

25 See also Horgan & Tienson (1988), who suggest that the complexity of routine decisions in a basketball game requires connectionist processing of structured representations rather than classical rule-based processing; Lloyd (1991), who argues that connectionist models work well for unconscious processes; and McClelland, Rumelhart, & Hinton (1986), who propose plausible connectionist models for many cognitive processes. Contrast Pylyshyn (1980: 111, 119), who claims that *all* cognition is "classically" computed.

Classical cognitive science says that cognition is rule governed symbol manipulation. . . . Connectionism says that thinking is activity in a neural network – taking the brain to be such a network. (Tienson 1991: 2)

The claim that thoughts have constituents does not entail that they are symbols (Chapter 20), and says nothing about how they are "manipulated." It is fully compatible with the claim that thinking is activity in a neural network.

I am aware of no other evidence that favors any form of the correlation hypothesis. We could conclusively show, for example, that Vladimir's playing C, playing E, and playing G were not part of Horowitz's playing the C-E-G triad if we determined that the first events took place in New York while the latter took place in Moscow, or that the Vladimir playing in New York was not Horowitz. I see no such evidence that my now thinking the thought that today is Tuesday or Wednesday does not consist of my now thinking the thought that today is Tuesday, the thought that today is Wednesday, and the concept of disjunction. There seems to be no good reason to maintain that the relationship between the words "Today is Tuesday" and the sentence "Today is Tuesday or Wednesday" is the part-whole relationship, while the relation between the thought that today is Tuesday and the thought that today is Tuesday or Wednesday is not.

§14.3 SUBPROPOSITIONAL CONSTITUENTS

In our defense of the constituency thesis so far, we have focused on establishing that *there are such things as thought-parts*. While disputed, this is not the most contentious claim that we need to establish. It is widely maintained that even if compound thoughts are composed of other thoughts, simple thoughts are indivisible – that they are "minimal units of mental action" and "syntactically unstructured."[26] On this view, the constituent structure of thoughts would resemble that of the formulas of propositional logic, in which the atomic constants are propositional, rather than

26 Lance & O'Leary-Hawthorne 1997: 357 and Peacocke 1983: 211, respectively. Following Geach 1957a: Chapter 14, Evans (1982: 101) says, "I should prefer to explain the sense in which thoughts are structured, not in terms of their being composed of several distinct *elements*, but in terms of their being a complex of the exercise of several distinct conceptual *abilities*. Thus someone who thinks that John is happy and that Harry is happy exercises on two occasions the conceptual ability which we call 'possessing the concept of happiness'" (cf. Peacocke 1983: 62–3, 209–10). But since Evans holds that *thoughts are complexes of exercises*, he must hold that exercises are the elements of thoughts, despite his mystifying wish not to hold this. The "possession" of concepts will be discussed in Chapter 16. Contrast Frege 1919: 254.

that of quantification theory (with its individual constants, predicate constants, and quantifiers) or of English (with its nouns, verbs, and adjectives). Simple thoughts are defined as those not composed of other thoughts. If *all* thoughts have constituents, then simple thoughts must have constituents that are not thoughts. If simple thoughts have such constituents, then all thoughts do. To complete our defense of the constituency thesis, therefore, we must establish that *thoughts have constituents that are not themselves thoughts.* We will describe such components as *subpropositional* and call them *ideas*.

The previous section assumed that "the idea of Tuesday," "the idea of John," "the idea of loving," and so on, are regular referential phrases denoting mental events that are not thoughts. Given this assumption, we argued that their referents are better construed as constituents of various thoughts than as a mere correlates. The constituency hypothesis is simpler, and provides a better explanation of the introspective integration of thoughts, the similarity, co-occurrence, systematic, and referential relations among ideas, the systematicity and productivity of thought, the compositionality of linguistic meaning, and the phrase-structure syntax of natural language sentences. We also observed that the idea of Tuesday could not be treated as a genus of which the thought that today is Tuesday is a species. We did not, however, explicitly consider the possibility that "the idea of Tuesday" denotes a *feature* of thoughts rather than a subpropositional component. We did argue that the idea that today is Tuesday could not be a feature of the thought that today is Tuesday or Wednesday. This idea is propositional, however. While thoughts are obviously events, it is not so clear that subpropositional ideas are events. So the opponent of subpropositional thought-parts could perhaps insist that subpropositional ideas are features rather than components of thoughts. However, given the similarities between the idea of Tuesday and the idea that today is Tuesday, the fact that the latter is not a thought-feature provides some reason to think that the former is not a thought-feature either. By postulating that ideas fall into two distinct ontological categories, the feature theory of subpropositional ideas is saddled with an ad hoc complexity. The constituency thesis is consequently simpler in one respect.

Further complexity is necessitated by this initial difference. If the feature theory is accepted for putative subpropositional components, we will have to revise Definition 15.1 to allow that ideas can be features as well as parts of thoughts, or else give up singular terms like "the idea of Tuesday." The revision will not be easy, because we cannot just say that any feature of a thought is an idea. We know of many features of thoughts that are

not ideas, such as their temporal, causal, and ontological properties. Even some intrinsic cognitive features, such as their having particular contents, and representing certain things, are not ideas. It is my belief that on the feature theory, we could not define "idea" in terms of "thought," and would have to introduce "idea" as an additional primitive term.

We would also be forced to give nonuniform analyses of expressions that appear to have the same logical form, and are treated uniformly by the constituency thesis. For example, on both theories "The idea that today is Tuesday occurred to S" can be analyzed straightforwardly as asserting that the idea referred to by the subject term "the idea that today is Tuesday" has the property expressed by the predicate "occurred to S." But on the feature theory, "The idea of Tuesday occurred to S" cannot be given a similar analysis because "the idea of Tuesday" is not construed as referring to an event. Since the idea of Tuesday never seems to occur to us unless some thought about Tuesday occurs to us (§14.1), the feature theorist might analyze "The idea of Tuesday occurred to S" as saying that some thought with the feature associated with "Tuesday" has the property expressed by the predicate "occurred to S." Similarly, whereas both theories can say that two thoughts are associated if one tends to cause the other to occur in a certain way (Chapter 18), the feature theory cannot say the same for the ideas of salt and pepper. The best the feature theory can do is to give an unnatural interpretation of "The idea of salt tends to cause the idea of pepper to occur," on which it means that thoughts with the feature associated with "salt" tend to cause thoughts with the feature associated with "pepper." It is unclear how a sentence like "In 'Mary loathes salt,' Jack used 'salt' to express the idea of salt" would be analyzed on the feature theory. It clearly will not do to analyze it as meaning that Jack used "salt" to express *some* thought with the feature associated with "salt." For Jack expressed a particular thought, and he used a sentence, not just the word "salt," to do it. These complications are minor, but they are ad hoc and completely unnecessary on the constituency thesis. If the feature theory had advantages over the constituency thesis, we might live with these wrinkles. But there are none.

Consider the argument from co-occurrence. One piece of evidence to be explained is that thoughts cannot occur to us unless certain subpropositional ideas occur to us. It is evident that if the thought that today is Tuesday occurs to us, or the thought that today is Tuesday or Wednesday, or the thought that today is Tuesday if Sally is receiving her piano lesson, then the idea of Tuesday must also occur to us. For we cannot think these thoughts without thinking of Tuesday. This is easily explained on the

hypothesis that the thought that today is Tuesday and the others have a common constituent, which is the referent of "the idea of Tuesday." The feature theorist can account for the fact that "The idea of Tuesday is occurring to S" is true in all of these cases by postulating that the thought that today is Tuesday, the thought that today is Tuesday or Wednesday, and so forth all have a common feature, which is associated with the word "Tuesday." But the feature theorist cannot explain why these thoughts have a common feature. The constituency theorist says that they have a common feature because they have a common component. The constituency theory thus has more explanatory power as well as being simpler. It might be thought that the constituency theory has no advantage on this score, since something is unexplained on both theories. But when the constituents of something are essential to its identity, they do not need to be explained. We cannot ask "Why does the thought that today is Tuesday contain the idea of Tuesday?" any more than we can ask "Why does H_2O contain hydrogen?"

Consider next the argument from similarity data. It is evident that the thought that today is Tuesday is similar in some way to the thought that yesterday was Tuesday, the thought that Tuesday comes after Sunday, and so on. It is also clear that the constituency hypothesis provides a natural explanation: all of these thoughts contain a particular component, called "the idea of Tuesday." The feature theory can also account for the similarities, of course. It is true even on the constituency thesis that these thoughts have a common feature. But the explanation provided by the feature theory is very shallow. "Two things are similar because they have common features" is nearly vacuous as an explanation. Explanations of similarity in terms of common constituents are much more illuminating. So again, the constituency thesis has more explanatory power.

Any theory of simple thoughts must account for the systematic similarities and differences among thoughts like the following.

Ljm The thought that John loves Mary.
Lmj The thought that Mary loves John.

The theory must also account for the fact that the relationships between these thoughts are similar to the relationships between the sentences that express them:

(1) John loves Mary.
(2) Mary loves John.

On the constituency thesis, the explanation is straightforward. The two thoughts are both composed of the same three ideas expressed by the three words occurring in both sentences. But the ideas occur in different relationships in the thoughts, just as the words occur in different relationships in the sentences. The fact that the structure of the thought expressed by a sentence is related in this way to the structure of the sentence makes a compositional semantics possible. The fact that simple thoughts have constituent structures of this sort explains why sentences of all natural languages have a phrase-structure syntax.

On the feature theory, the thoughts Ljm and Lmj have no components. The thoughts are similar because they have three features in common, associated with the words "John," "Mary," and "loves." Call these features J, M, and L. Since the thoughts do differ, the feature theory must also postulate that the thoughts have a fourth feature, which is different. Let δ_1 be the distinguishing feature of Ljm and δ_2 the distinguishing feature of Lmj. It seems that δ_1 must relate J, M, and L in a certain way, and that δ_2 must relate J, M, and L in a different way. It similarly seems that thought Hjm must have features J, M, H, and δ_1, while thought Hmj has features J, M, H, and δ_2.

Hjm The thought that John hates Mary.
Hjm The thought that Mary hates John.

Nagging questions arise: What features could δ_1 and δ_2 be, and why do certain thoughts have them and not others? Why is it that a person able to think a thought with feature δ_1 is also able to think a thought with feature δ_2? To account for the relationships among simple thoughts, the feature theory must postulate that thoughts have the sorts of relations among their features that would be expected if thoughts had constituent structures. But on the feature theory, simple thoughts have no internal structure. So how can they have the requisite relationships among their features? In a number of ways, then, the feature theory is more complex than the constituency thesis, and leaves more unexplained.

We will conclude with a more direct argument that thoughts have subpropositional components. Consider the thoughts expressed by the following sentences:

(3) It is not the case that John is tall.
(4) It is probable that John is tall.
(5) It is too bad that John is tall.

The arguments of section §14.2 show that these thoughts have the thought that John is tall as a component. Since the thought that John is tall is not identical to any of these thoughts, it must be a *proper* part of them. That can be the case only if they have some *other* components. They do not appear to have any other thoughts as components. So the extra components must be subpropositional ideas. The feature theorist would have to maintain that the thought that it is too bad that John is tall and the thought that John is tall are *separate* thoughts, which have most but not all features in common, and which are such that the first never occurs without the second. But then the arguments against the correlation thesis become arguments against the feature theory. The feature theory does not explain why the three thoughts all occur with the thought John is tall, why they are similar, or why there appears to be a unified thought process in each case.

Consider next the following expressions:

(6) John kissed Mary.
(7) John kissed Mary Tuesday.
(8) Tuesday.

Reading sentences (6) and (7) makes us conceive complete thoughts. The arguments of §14.2 show that the thought that (6) makes us conceive is a constituent of the thought that (7) makes us conceive. The argument of the previous paragraph shows that the thought that (7) makes us conceive has an additional subpropositional component beyond the component that (6) makes us conceive. Now, reading word (8) also makes us conceive something. Indeed, it makes us think of Tuesday. It is natural to suggest that the idea that (8) makes us conceive is at least one of the extra constituents that makes the idea expressed by (7) more complex than that expressed by (6). On the feature theory, reading (8) either does not make us conceive any idea, or it makes us conceive a thought with a distinctive feature. Either way, the feature theory cannot maintain that (8) expresses the extra component differentiating (7) from (6).

Consider finally this sequence of expressions:

(9) John
(10) John got
(11) John got nothing.

Reading all three of these expressions makes us conceive something. By §14.2, what (10) makes us conceive contains what (9) makes us conceive,

and is contained in what (11) makes us conceive. The reading of "got" in (10) makes something occur to us that does not occur to us when we read (9), but which does occur to us when we read (11). What (10) makes us conceive, furthermore, is not a complete thought, in contrast to what (11) makes us conceive. It is introspectively evident that (9) makes us *think of John*, that the reading of "got" makes us think of getting, and so on. The hypothesis that the words in (11) are associated with properties of the thought expressed by (11) does not account for any of these facts. The hypothesis that the words express distinct subpropositional thought-parts accounts for all of them. To account fully for the compositionality of semantics, we must postulate that thoughts have subpropositional constituents.

§14.4 THE GENERAL CONSTITUENCY THESIS

We noted in §12.1 that the word "thought" is ambiguous, denoting either relational *objects* of thought (such as propositions) or *acts* of thinking. Postulate 14.1 concerns relational objects of thought, and asserts that they are complex. Are acts of thought complex too? The answer is yes. For *the thinking relation is simply the inverse of the occurrence relation* (Definition 12.1). The thought that it will rain is a relational object of thought: something that you and I both may think, something that may be true or false, and so on. For us to think that thought is simply for that thought to be occurring to us. To ask whether thinking a complex thought is itself complex is to ask whether the occurrence of a complex event is itself complex. If Definition 12.1 is questioned, the arguments for the complexity of objects of thought can be repeated, mutatis mutandis, for acts of thought. For example, the fact that one cannot think the thought that today is Tuesday or Wednesday without thinking the thought that today is Tuesday is best explained by the hypothesis that the latter act is part of the former.

Matters are quite different with the other propositional attitudes. Consider the act of belief, another meaning of "thought" noted in §12.1. Being propositions, the relational *objects* of belief are complex. But the "act" of believing a proposition does not share the complexity of that proposition. First, note that the belief relation is not simply the inverse of the occurrence relation. As emphasized in §12.1, a proposition can be believed without being occurrent, and can be occurrent without being believed. More importantly, the parts of a believed proposition need not be believed, as disjunctions and conditionals make clear. The

components of atomic propositions are not even capable of being believed. Conversely, a proposition whose parts are believed need not itself be believed; witness negations and certain conditionals.

Fodor (1987: 135–54) held that propositional attitudes *in general* are complex, which I call the *general constituency thesis*. As he put it, "the syntactic structure of mental states mirrors the semantic relations among their intentional objects" (1987: 138).[27] Unfortunately, Fodor selected an unrepresentative set of intentional objects (conjunctions) and propositional attitudes (intending and believing) to illustrate his thesis. It is undeniable that someone who believes a conjunction believes its components; it is at least plausible that the same goes for intending. Given these correlations, a hypothetical induction like that sketched earlier is initially plausible. But, as Fodor (1990a: 18) later realized, the conclusion that the state of believing P&Q has as constituents the state of believing P and the state of believing Q is undermined by the fact that similar conclusions fail demonstrably for other attitudes (wanting, for example) and other logical compounds (such as disjunctions). Since one can believe the proposition that today is Tuesday or Wednesday without believing the proposition that today is Wednesday, the act of believing the former proposition cannot contain the act of believing the latter as a part, even though the objects of the two beliefs are so related. Similarly, wanting it to be the case that one wins both the new car and the crock pot does not imply wanting it to be the case that one wins the crock pot, which shows that the desire for a conjunction does not contain the desires for its conjuncts.

Fodor offered several arguments for the general constituency thesis. One was a variant on the simplicity argument sketched here for objects of thought. Other things equal, it is simpler to explain a complex effect in terms of the sum of the causes of the individual effects than in terms of some additional cause co-occurrent with that sum. But this sort of argument can be deployed only for a very limited sample of propositional attitudes and objects, such as the believing of conjunctions. Another argument that Fodor advanced was that it is impossible to explain the

27 Cf. Loar 1981: 126, expounding a suggestion of Field (1978); Barwise & Perry 1983: 223, 242; Maloney 1989: 4–7; Crimmins & Perry 1989; Horwich 1998a: 44. See also Geach 1957a: 49: "analysing judgments in terms of 'objectives' is a lazy analysis; when judging is treated as a simple two-termed relation, the complexity of the judgment is just transferred in its entirety to that which is judged, the 'objective'"; and Harman (1973: §4.1; 1978: 63–4), who claims that beliefs and other mental representations have "logical structure."

systematicity or productivity of thought without postulating that thinking is complex. With that postulate, the explanation is simple:

> [H]aving a thought is being related to a structured array of representations; and, presumably, to have the thought that John loves Mary is ipso facto to have access to the same representations, and the same representational structures, that you need to have the thought that Mary loves John. So *of course* anybody who is in a position to have the one of these thoughts is ipso facto in a position to have the other. (Fodor 1987: 151)[28]

This is an admirable argument, I believe, but only for the act of thinking. First, I do not think that other propositional attitudes are as systematic as thought is. For example, anyone who can think "John is taller than Mary" is surely capable of thinking "John is taller than John." But it is doubtful that anyone is psychologically capable of believing that John is taller than John. Similarly, someone who can intend it to be true that John is prevented from carrying out every one of his intentions can surely think the thought that he himself is prevented from carrying out his intentions. But it is not clear that anyone is capable of intending that he be prevented from carrying out every one of his own intentions. Second, to the extent that other propositional attitudes are systematic, there is an alternative explanation. *Believing and intending are systematic,* I believe, *to the extent (i) that thinking is systematic and (ii) that anything S is capable of thinking S is psychologically capable of believing and intending.* The same goes for all other propositional attitudes. We do not need to debate whether this explanation is as simple as Fodor's, because we have independent proof that the constituency hypothesis is untenable for believing and intending.[29] Fodor later proposed a third hypothesis, that propositional attitudes are systematic because they are relations to *symbols*, which of course are systematic. This

28 See also Evans 1982: 103–4; Fodor 1990a: 16–17; Peacocke 1992: §2.1; Crimmins 1992: 76; and Carston 1994: 4610.

29 Fodor had a third argument, which was too sketchy for me to get a handle on. He implied that the constituency hypothesis provides an explanation for mental *processes*. His idea seemed to be that transitions from one propositional attitude to another can be explained by supposing that the subject "moves" parts of the attitude. Unfortunately, I do not know how thoughts or ideas can be moved, let alone beliefs or desires. Moreover, let us assume that believing P & Q has believing P and believing Q as components. I do not see any operation applicable to the assumed components here that would shed any light on the process of moving inferentially from believing P & Q to believing Q & P. Note that even without such explanations, it remains true that mental representations "constitute domains over which *mental processes* are defined" (1987: 145). For the inference from P & Q to Q & P cannot be described except in terms of the objects of the initial and terminal beliefs, namely P & Q and Q & P, and these are mental representations, viz., thoughts.

view, called the language of thought hypothesis, will be criticized in Chapter 20.

Perry and his coauthors define ideas and notions as parts of *beliefs*.[30] This would coincide with Definition 15.1 if it were interpreted as meaning that ideas are parts of objects of belief. For propositions are the objects of belief, and propositions are thoughts of a certain kind. Indeed, since "thought" is often treated as a synonym of "belief," Perry's definition may seem to differ only superficially from mine. This impression may be reinforced by the fact that Perry deploys the terms "idea" and "notion" in much the same way that I do. Thus Perry speaks, as I do, of "forming," "having," and "expressing" ideas. He uses expressions like "the idea of Edward," and assumes that the word "royal" expresses the idea of being royal. But by "belief," Perry means the "act" of believing – the state we enter when we acquire the belief, and in which we remain until we no longer have the belief. My belief that $2 + 2 = 4$ is a state I have been in continuously since I was at least four years old. Given that "thought" in the sense I intend is markedly different from "act of belief," there are substantial differences between the two definitions. (1) While Perry's presupposition that beliefs are complex may seem every bit as justified as my assumption that thoughts are complex, the evidence for the complexity of thought is stronger. Indeed, I believe that while there is plenty of evidence that the objects of belief are complex, there is no evidence that the state of belief is. So it is not as clear that there are any ideas, as Perry defines them. It does not seem possible to maintain that the "act" of believing a proposition contains the same ideas that the proposition contains. Of course, no one can have the belief that Napoleon is French without having the concepts of Napoleon and of being French, and the belief cannot be occurrent unless the ideas are. However, we can explain why believing something implies having certain concepts without postulating that the state of believing contains concepts. A proposition is a concept, specifically, a thought; S cannot believe a proposition unless S has that concept; finally, S has a complex concept only if S has all of the concepts contained in it, because a concept cannot occur to A unless all of the components do. I do not know of any data unexplained by this account. The notion of having concepts will be examined in Chapter 16.

30 See Barwise & Perry 1983: 242; Crimmins & Perry 1989; and Crimmins 1992. Cf. Fodor 1998a: 6. A seperate difference is that Crimmins and Perry define ideas as *token*s rather than types, which implies that people can never share ideas, or think the same thing twice. It also implies that people can have more than one idea of a cat.

Let us assume for the remainder of this comparison that beliefs do have parts. (2) Perry would also need to explain how belief-parts are related to desire-parts. It would seem that the belief that John won and the desire that John won involve the same ideas (of John, and of having won) related to each other in the same way (as subject-concept and predicate-concept in a singular proposition). But can a belief-part be part of a desire? If the belief and the desire consist of the same things related to each other in the same ways, how can the belief and the desire be distinct states? If the belief and the desire have different constituents, how can they involve the same ideas? Must we deny that words in nonindicative sentences express ideas? On my view, these questions do not arise: the belief and the desire involve the same ideas because they have the same propositional object composed of those ideas; the states differ because they are different relations to the same object.

The next two differences are more important. (3) Unlike thinking, believing is a dispositional state rather than an event or occurrence (Chapter 12). The acquisition or loss of a belief is an event, of course. But to say that John believes that $2 + 2 = 4$ is not to say that he is active in any way, or that anything is happening to him, or that anything is going on in his mind at the moment. Since beliefs are not events, the parts of a belief cannot not be events. So ideas, on the Perry definition, are not the sorts of things that could occur to us. We could not, therefore, speak of the "association" of ideas in anything like its traditional sense (see Chapter 18). Furthermore, we would have to give up Definition 12.2, according to which S is thinking of Φ iff the idea of Φ is occurring to S. Something other than the mere presence of the idea of the number 2 would have to account for what it is for us to be thinking of that number. (4) Suppose that while making up a sequel to the *Star Wars* trilogy with my son, I say and mean "Luke Skywalker then turned to the dark side of the Force." I am expressing a thought, but not a belief. Hence I am using "Luke Skywalker" and "dark side" to express thought-parts but not belief-parts. So if we adopted Perry's definition, we would have to either redefine expression in some way, or give up the thesis that what a speaker means by a word is determined by what idea he is using it to express. In sum, it is not clear that there is anything satisfying Perry's definition of "idea"; his definition does not accord as well as Definition 15.1 with the traditional and conventional use of the term "idea"; and most important, if any entities do satisfy Perry's definition, they cannot play the role assigned to thought-parts in my theory.

We are using the terms "part," "constituent," and "contain" in their *mereological* sense. The word "containment" has often been used in logic, however, to signify that a conclusion follows logically from a set of premises. Thus the conclusion C_1 that John is a mathematician is said to be contained in the premise P_1 that John is a mathematician and a philosopher. Since C_1 is also a part or constituent of P_1, it is natural to hypothesize that logical containment is, or is explained by, mereological containment.[31] But a moment's reflection will deflate any such speculation. The conclusion C_2 that John is a mathematician or a philosopher is logically contained in the premise P_2 that John is a mathematician, but not conversely. The reverse holds for mereological containment: P_2 is a proper part of C_2. The fact that C_2 is not a part of P_2 is shown by the fact that C_2 is more complex, and can occur to someone without P_2 occurring to him. To use Frege's apt metaphor, C_2 is not contained in P_2 in the way that beams are contained in a house, but more in the way that plants are contained in their seeds.

The logical principle that the truth of a conjunction entails the truth of its conjuncts is not directly relevant to the question of the components of a conjunction. For this principle does not entail that people cannot *believe* a conjunction without believing its conjuncts, if by "cannot" we mean "cannot *psychologically*" rather than "cannot *rationally*." Still less does the logical principle entail that people cannot *think* a conjunction without thinking its conjuncts. Furthermore, the same co-occurrence relations obtain between propositions that are not related logically in the same way. In the example used earlier, it was observed that a *disjunction* cannot occur to a subject unless its disjuncts do. This is true even though the truth of a disjunction does not entail the truth of its components, and even though people can believe a disjunction without believing its components.[32]

31 Cf. Leibniz as discussed in Mates 1986: 60–1, 87. Contrast Frege 1884: 101; Fodor 1981: 299–301; 1998a: 88– 9, 108–12; Katz 1986b: Chapter 5; 1990 Fodor & Lepore 1992: 49.

32 A parallel observation can be made concerning the issue of lexical decomposition. Fodor and others have proposed accounting for the semantic relationships between "kill" and "die" in terms of "meaning postulates," such as that "A killed B" entails "B died," rather than in terms of "compositional postulates," such as that the concept of dying is a constituent of the concept of killing (see Fodor 1981: Chapter 10; Chierchia & McConnell-Ginet 1990: 350–6). One fact explained by the latter postulate but not the former is that people cannot think of killing without thinking of death. Meaning postulates concern entailment relations, not co-occurrence relations. I hasten to add that the constituency thesis advocated in the text does not entail the lexical decomposition thesis (cf. *Nondescriptive Meaning and Reference*).

Consider the following passage from the *Port-Royal Logic*.

> The comprehension of an idea is the constituent parts which make up the idea, none of which can be removed without destroying the idea. For example, the idea of a triangle is made up of the idea of having three sides, the idea of having three angles, and the idea of having angles whose sum is equal to two right angles, and so on. (Arnauld 1662: 51)[33]

The question is, was Arnauld using "constituent part" in its mereological sense? His statements that the parts of an idea "make up" the idea, and that the parts cannot be "removed" without "destroying" the idea suggest that the answer is yes. But in that case, his claim that the idea of having angles equaling two right angles is a constituent of the idea of a triangle is demonstrably false. We need only observe that it is altogether possible for the idea of a triangle to occur to someone without the idea of having angles whose sum equals two right angles occurring to him. Indeed, that will surely be the case until the person proves or otherwise learns that the angles of a triangle add up to 180°. When seven years old, my son had been conceiving of triangles for years. But he did not even then have the concept of 180°. We may grant that a figure cannot *be* a triangle without having angles adding up to 180°. But that would show only that the idea of the latter is *logically* contained in the idea of the former (assuming that we have extended the notion of logical containment from propositions to ideas in the obvious way).

Consider next Kant's definition of the analytic-synthetic distinction.

> In all [affirmative] judgments in which the relation of a subject to the predicate is thought . . . , this relation is possible in two different ways. Either the predicate B belongs to the subject A, as something which is (covertly) contained in this concept A; or B lies outside the concept A, although it does indeed stand in connection with it. In the one case I entitle the judgment analytic, in the other synthetic. . . . The former, as adding nothing through the predicate to the concept of the subject, but merely breaking it up into those constituent concepts that have all along been thought in it, although confusedly, can also be entitled explicative. The latter, on the other hand, add to the concept of the subject a predicate which has not been in any wise thought in it, and which no analysis could possibly extract from it; and they may therefore be entitled ampliative. (Kant 1787: 48)

It would appear from this passage that Kant is taking "contains" and its relatives in the mereological sense. But if he is, then he would have to count "John's mother is John" as analytic. For the predicate-concept

33 Cf. Leibniz 1709: 2.6; Descartes 1641b: 54; Mates 1986: §5.2.

"John" is a constituent of the subject-concept "John's mother." The concept "ocean" is part of the concept "ocean drive," so "An ocean drive is an ocean" would have to count as analytic. Since the propositions in question are patently false, it is more charitable to assume that for Kant containment is, or at least implies, logical containment.[34]

The most we can say about the relation between logical and mereological containment is this: to the extent that logical entailment can be explained in terms of propositional structure, it can be explained in terms of mereological containment. But it is never the mere mereological containment of Q in P that explains why P entails Q, but rather P's containing Q in a certain way (e.g., as a conjunct), or Q's containing P in a certain way (e.g., as a disjunct), and so on.

The conflation of mereological and logical containment is just one of the many elementary confusions that have obscured the theory of ideas for centuries. Others will be reviewed in Chapter 19.

34 Cf. Katz (1964b: 531; 1966: 188–95; 1974: 305), whose definition of analyticity would seem to have the same difficulty. See also Linsky 1972; Mates 1986: 61; the "definition theory" as defined in Fodor 1998a: 88, 108; and Murphy & Medin 1995: 445.

15

Ideas or Concepts

Chapter 14 introduced the thesis that ideas or concepts are thought-parts, and sought to clarify and establish the underlying assumption that thoughts have constituents. This chapter will formally define "idea," and set out the basic properties of concepts. We will introduce the distinction between atomic and complex concepts, and analyze what it is to conceive concepts. We will discuss the content of concepts, and the fact that concepts represent objects. And we will distinguish the intentional content of ideas from their extension or objective reference.

§15.1 FORMAL DEFINITION

The terms "idea" and "concept" are ambiguous in English, and have been used with many different meanings in philosophy and psychology. In particular, they have been used to mean universals (e.g., Husserl), sensory images (Hume), objects of thought (Descartes), contents (Burge), senses (Katz), conceptions or belief systems (many contemporary psychologists), and mental representations generally (Brentano). We will use "idea" and "concept" exclusively for *thought-parts*,[1] distinguishing them from

1 While I believe that this definition best fits the role traditionally assigned to ideas in ideational theories of meaning, it has been seldom suggested and rarely adopted consistently. Ockham provides perhaps the clearest precedent. While he first offered the Cartesian definition (*Ordinatio*: 46), Ockham later abandoned it for that of a proposition part (*Summa Logicae I:* §1). Leibniz followed Ockham in taking concepts to be the components of propositions (1709: 7.1; see also Mates 1986: 58; Bealer 1998: 267); but Leibniz also offered the Cartesian definition (1709: 2.1). The claim that concepts should be defined as thought parts is at least a natural interpretation of some of Frege's remarks: "I do not begin with concepts and put them together to form a thought or judgment: I come by the parts of a thought by analysis of the thought" (see 1979: 253; see also 16–17). However, Frege seems to infer from

universals and other objects, images, and conceptions (see Chapter 19) as well as from contents (§15.6) and senses (§21.1). An idea, on my conception, is only one kind of mental representation.[2] Whereas ideas or concepts are wordlike mental representations, images are picturelike, and conceptions are theorylike.

Even when "idea" and "concept" both denote thought-parts, they have somewhat different patterns of usage in standard English. The terms are equally natural when discussing expression or containment. They are interchangeable when discussing the contents and objects of thought-parts, and what they represent. But "idea" is more natural as a cognate of "thought," and "concept" as a cognate of "proposition." "Idea" is idiomatic when discussing the association of ideas. "Concept" is idiomatic when discussing the acquisition, possession, and conceiving of concepts. "Have" has a different meaning when followed by "idea"

this that the content of a concept is somehow derived from the content of the thoughts it is part of. And after saying that thoughts have to have parts elsewhere, he remarks that "we really talk figuratively when we transfer the relation of whole and part to thoughts" (1923: 55). Geach (1957a: 12–14) held that "a judgment is the exercise of a number of concepts," but defined concepts as abilities to use words. Hebb (1966: 91) suggested that "idea" could be defined as "elementary component of thought," but went on to suggest replacing the term with "mediating process." Burge (1979a: 425–6) defines concepts as thought parts, but simultaneously offers the Cartesian definition (§19.1, this volume). Burge (1979b: 537) elsewhere takes concepts to be parts of thought *contents*, as does Boghossian (1998b: 257). Peacocke (1983: 56, 106; 1986: 1; 1992: 2, 99–100) too defines concepts as thought components while identifying thoughts with contents; but the seriousness of his mereological commitment is doubtful, since (1) he also goes on to say that "the talk of composition is just a vivid way of encoding more or less complex intrinsic features of the patterns of canonical grounds and commitments which ultimately individuate the thought" (1986: 63; cf. 1986: 114; 1992: 118–9; contrast §14.2 here); (2) he is equally willing to say that objects, like people, can be constituents of thoughts; and (3) he denies elsewhere that thoughts are syntactically structured (1983: 206–18). E. E. Smith (1988) and Jackendoff (1989) take concepts to be thoughts or thought components, but also seem to treat them as conceptions (things known or containing knowledge, "information structures"). Rey (1994) defines concepts as constituents of "propositions," but then says that they could be words, images, definitions, or prototypes. Barwise & Perry (1983: 242), Crimmins & Perry (1989), Perry (1990), Crimmins (1992: Chapter 3), and Horwich (1998a: 44) introduce ideas (and notions) as components of *beliefs* (contrast §14.4). Finally, see Fodor's notion of "symbols in the language of thought," discussed in Chapter 20, which he says are "often called 'ideas' in the older literature." Fodor now defines concepts as constituents of beliefs and other mental states (1998a: 6), and takes them (impossibly) to be both particulars and types (1998a: 22, 28). Fodor shuns the word "idea" because so many people take ideas to be images. It is unclear whether Fodor considers concepts to be mental representations, or to be "modes of presentation" that differentiate mental representations with the same objective reference (1998a: 21–2).

2 Contrast Yolton 1956: 96; Arnauld 1662; Mill 1843: 4.2.1; 1865: 419; Brentano 1874: 38–9, 198; Meinong 1910: xii; and Warren 1921: 5.

than when followed by "concept." The different connotations and usage patterns of "idea" and "concept" may lead one to suspect that the terms have different denotations in ordinary English, but I have not seen any solid evidence to that effect. We will treat "idea" and "concept" as stylistic variants.

As noted before, there is a tendency to reserve the terms "idea" and "concept" for *proper* parts of propositions, those that are not themselves propositions. But we will allow that it is proper to refer to the idea or concept that man evolved from the apes, a proposition. There is a marked tendency in philosophy and psychology to use "concept" only for *general* as opposed to *singular* concepts, to speak, for example, of the concept of man but not of the concept of Socrates.[3] Hence concepts are often identified as mental representations of *universals* or *categories*, defined by their role in *classification*. But the proposition that Socrates is a man has at least two parts, and we will call both "concepts." Since we can think about Socrates, we have to have a concept of Socrates.[4] This usage does not prevent us from recognizing that general concepts are different from singular concepts in important ways. But we must equally well recognize the similarities. Both are involved in applying knowledge to perceived objects, for example. We can perceive an object to be Bill Clinton just as surely as we can perceive it to be a man. And when we do either, we apply our knowledge, whether of Clinton or of men. *Identification* is just as important a cognitive process as classification. *Recognition* may involve either process. A particularly important similarity in this book is the fact that both singular and general concepts can be expressed by words, the former by singular terms (e.g., "Bill Clinton"), the latter by general terms (e.g., "man").

Our standard method of referring to an idea, whether a thought or thought-part, is by displaying an expression μ that expresses that idea in the language we are using (on the occasion of its use). Using *oratio recta*, we typically display the expression by either placing it in quotation marks or italicizing it. Thus *the idea "cats"* refers to the idea expressed by "cats" in English, just as *the thought "Tabbies are cats"* refers to the thought expressed by "Tabbies are cats" (§7.6). It follows that idea μ is identical to idea ν

3 For examples of both tendencies, see Frege 1892a; Stout 1899: 395–6; Joseph 1916: 22; Titchener 1914: 219; Humphrey 1951: 288; Kneale & Kneale 1962: 603; Boehner 1964: xxvii; Vendler 1972: 69–77, 132–135; Smith & Medin 1981; Peacocke 1983; 1986; Bealer 1998: 267. For the first, see Reid 1785: 384–5; Peacocke 1992; Fitch 1993: 471; Margolis & Laurence 1999: 4.

4 This inference will play a key role in my *Nondescriptive Meaning and Reference*.

iff "μ" and "ν" have the same meaning (cf. §13.6). Using *oratio obliqua*, "the idea of ϕ" denotes the idea expressed by the pronoun-free object nominal "ϕ" (Definition 7.9), just as "the idea that ρ" refers to the idea expressed by the pronoun-free declarative sentence "ρ" (Definitions 7.7, 13.2). Hence the idea of cats is the idea "cats," and the idea that everything is relative is the idea "Everything is relative."

Expressions of the form "the idea of Φ" are singular terms denoting event-*types*. There is just one idea of a cat, with countless tokens.[5] Two people may have the same idea, but not the same token of an idea. Also, there are plenty of ideas that no one has ever had. The general terms "idea" and "concept" also denote *types*; their extension includes the idea of 9, the idea of being odd, and so on. Unlike the term "thought," the terms "idea" and "concept" have no token sense in standard English (cf. §12.3). While there have been many conceptions of Earth, there is only one concept "Earth." And there is no sense in which having more people conceive of the Earth increases the number of concepts that there have been. There are idea-tokens, of course, but they are not denoted by the word "idea" in conventional English.

As we are using the term, concepts are wordlike mental representations. But concepts are not words. Words contain letters or speech sounds, and belong to particular languages. Concepts do not. Concepts are components of thoughts, words are components of sentences. Words express concepts, just as sentences express thoughts. We use words, but not concepts. Even though our usage conforms to standard English usage, many philosophers, psychologists, and linguists have referred to words as concepts, either by inattention or by stipulation. Even those who distinguish concepts from words sometimes apply terms to concepts that are more naturally applied to words. Thus we are sometimes said to "learn," "use," and "apply" concepts, and they are often said to have "meaning."

Concepts cannot be defined as *belief-parts* unless we are talking about objects of belief rather than acts of believing.[6] The objects of belief are propositions, which we identified with thoughts in §13.3. So the claim

5 Contrast Crimmins & Perry 1989: 688; Crimmins 1992: Chapter 3; Gauker 1994: 29–30. Fodor (1998a: 23) and Margolis and Laurence (1999: 5–8) claim that concepts are *particulars*, and Perry (1990) claims that beliefs are complex particulars. If concepts were particulars, there would be just one concept of a cat, which could not occur to several different people at the same time, or to the same person at different times. This conflicts with their claim in other places that concepts are *public* entities, *types* that can be shared and *tokened* (Perry 1990: 21; Fodor 1998a: 28; Margolis & Laurence 1999: 76–7). Types are universals, not particulars.

6 Compare and contrast Barwise & Perry 1983: 242; Crimmins 1992: Chapter 3; and Horwich 1998a: 44.

that concepts are parts of beliefs in this sense is equivalent to the claim that they are parts of thoughts. But the propositional attitude of belief does not have parts, as we argued in §14.4. Thinking is a complex event. Believing is neither an event nor complex, although its objects are complex. Thought is the only propositional attitude whose instances have parts.

We argued at length in Chapter 14 that ideas are parts rather than features or correlates of thoughts. Before we can officially define ideas as thought-parts, however, we need to consider whether being a thought-part is sufficient as well as necessary for being an idea. While we established its converse, we did not argue for the thesis that *all thought-parts are ideas*. To my knowledge, this thesis has never been examined in the literature. There are no *known* thought-parts other than ideas. So we could propose an inductive argument: "All known thought-parts are ideas, therefore all are." But it is conceivable that thoughts have other sorts of parts as well. Complex ideas have other ideas as parts. Atomic ideas, by definition, do not have other ideas as parts (see §15.2). That leaves open the possibility that atomic ideas have parts other than ideas. If that is the case, then thoughts have parts that are not ideas. For any part of an idea is part of a thought, given that all ideas are thought-parts and given that constituency is transitive. The possibility that ideas have nonideational parts is more than a logical possibility, I believe. There is abundant support for the view that mental events are neurophysiological processes. There is a real possibility, then, that ideas are neural processes. But any neural process is bound to have spatial or temporal parts too small to be mental events themselves. The sentential analogy may help to clarify the possibility that some parts of thoughts are not ideas. For while every word is a part of some sentence, not every part of a sentence is a word. Since words are composed of speech sounds or letters, so are sentences. But individual letters and the sounds that they represent are not themselves words (with a few exceptions).

As long as there is a real possibility that thoughts have parts other than ideas, our definition of "idea" should specify what distinguishes the parts of thoughts that are properly called ideas from those that are not. I believe we can use the fact that *the qualitative, intentional, semantic, and attitudinal properties of a thought depend on the ideas composing it*. In any given position in a thought, different ideas can occur. The thought that Berlin is the capital of France is identical structurally to the thought that Paris is the capital of France, except that the former has the idea of Berlin where the latter has the idea of Paris. Because of this difference in component ideas, the thoughts contrast in at least four important respects. *Qualitative character*:

The difference in component ideas makes a difference in what it is like to think the thoughts. Thinking the thought that Berlin is the capital is a different conscious experience from thinking the thought that Paris is the capital. *Intentional content*: The difference in ideas makes a difference in what a person is thinking who thinks the thoughts. One thought is about Berlin, the other about Paris. One represents Berlin as the capital, the other represents Paris as the capital. *Semantic properties*: The difference in ideas makes a difference to the truth value of the thought. One thought is true, the other false. The thoughts have different entailments and incompatibilities. The contrast between the idea of Paris and the idea of the capital of France makes the thought that Paris is the capital of France a contingent truth, whereas the thought that Paris is Paris is a logical truth. *Attitudinal status*: Because of the difference in component ideas, people can have different propositional attitudes toward the thoughts. The proposition that Paris is the capital of France is something that I believe, while the proposition that Berlin is the capital is something that I disbelieve. That Berlin is the capital of France might well have been something that Kaiser Wilhelm desired, preferring that to Paris's being the capital. It appears that thoughts differing in one of these four respects differ in all of them.

Let us call a part that is capable of making a difference in the qualitative character, intentional content, semantic properties, or attitudinal status of a thought a *cognitive part*. I believe that any thought-part incapable of making such a difference would not properly be called an idea. If thoughts have any neural parts too small to be mental events themselves, for example, they could not make a difference to the qualitative character, intentional content, semantic properties, or attitudinal status of any thought. We will therefore define ideas as cognitive thought-parts.

15.1 **Definition:** *Ideas (concepts) are thoughts or cognitive parts thereof.*

The sentential analogy can be pressed into service again to help clarify the notion of a cognitive part. Some sentence parts make a difference in the semantic and grammatical properties of a sentence, but others do not. Those parts that do make such differences are the *syntactic constituents* of sentences. Words and letters are obviously syntactic constituents. The ink drops forming the letters are not syntactic constituents, because the sentence (token) would still have the same semantic and grammatical properties if it had toner particles in place of the ink drops. Given the formal similarities between sentence structure and thought structure, it is natural to use the term "syntax" in a generalized sense, in which the cognitive parts of a thought are its syntactic constituents (see §20.5).

Our basic primitive terms are "thought" and "occur." We have used them to introduce the more general term "idea," the relation "S is thinking T," and the singular terms "the thought that p" and "the idea of Φ." We have used the latter to introduce "S is thinking of Φ." By Definition 12.2, *S is thinking of Φ iff the idea of Φ is occurring to S.* This rule is one that would fail if "idea" were defined to mean either "image" or "conception." I can think of a chiliagon even though no image of one is occurring to me. I can think of the universe even though there is no single conception of the universe, and even though my own is too complex to occur to me all at once when I am thinking of the universe. The self-evidence of Definition 12.2 in standard English appears to have led associationist psychologists such as J. Mill (1829) into lapses of rigor. For even though they defined ideas as images, they still automatically inferred that the idea of something is occurring to a subject from the fact that he is thinking about it. The inference is obviously valid when ideas are defined as thought-parts, but not otherwise.

§15.2 ATOMIC IDEAS

I have tried to establish that some ideas have other ideas as parts. Such ideas are said to be *complex*. Ideas that do not have ideas as parts are *simple* or *atomic.*

15.2 **Definition:** *An idea is complex if it contains other ideas as parts; otherwise it is atomic (simple).*

A familiar type of reductio ad absurdum argument shows ideas cannot all be complex.

15.3 **Postulate:** *Some ideas are atomic.*

Some ideas must be atomic, because otherwise there would be an infinite regress or circle of ideas. There cannot be a circle, because "x is a (proper) part of y" is an order relation, transitive, asymmetric, and irreflexive. If there were a circle, then some idea would be a part of a part of a part . . . of a part of itself. By transitivity, it would be a part of itself, violating irreflexivity. There cannot be an infinite regress either. If there were, then the parts of any idea would have to have ideas as parts, and those ideas would have to have ideas as parts, and so on ad infinitum. Every thought and idea of even the least intelligent and mature person would have to be infinitely complex. But this is absurd, given that human experience and cognition are finite and limited.

Given the observation made in §14.1 that every thought is composed of at least two other ideas, it follows that thoughts are not simple ideas. This means that only subpropositional ideas are atomic. If a subpropositional idea is complex, then it is in principle *definable*, and definable in the strictest sense. If idea i is composed of ideas j and k, then there could at least be terms "I," "J," and "K" expressing those ideas in such a way that "I" means "JK." The qualification "in principle" is necessary because the languages available to us need not have the requisite words and constructions. When I say that a term is definable in the strictest sense, I mean that the definiens and definiendum are not only coextensive but also synonymous: they have the same meaning because they express the same idea.[7] It is plausible that "vixen" expresses a complex idea, because it is plausible that "vixen" means "female fox." That is, a vixen is definable as a female fox, and the definition "A vixen is a female fox" is not just true but analytic. A term or idea that is not defined in a given system is said to be *primitive*. Atomic ideas are primitive in an absolute sense: they cannot be analytically defined in any system. Psychologists refer to the thesis that all concepts are definable as the *classical theory*. The argument that some ideas must be simple has the same logic as the familiar argument proving that in any system of definitions, some terms must be undefinable or primitive. The thesis that not every word is definable holds for "semantic" or "analytic" definitions, in which the definiens is a complex expression synonymous with the definiendum.[8] If every word were definable, then there would be either an infinite regress of definitions, or a circle of definitions. Since there is only a finite number of simple words (those that do not contain other words as parts), there cannot be an infinite regress. There cannot be a circle of analytic definitions, given that the definiendum means what the definiens does, and that the meaning of the definiens is determined compositionally by the meanings of its components. If there were a circle, then the definiendum would have to be synonymous with some complex expression containing itself as a proper part. Then the definiendum would have to mean a proper part of what it means, which is an impossibility.

The expression theory of meaning leaves it completely open whether or not any words express atomic ideas. Atomic ideas must exist, but there is no argument that any of them must be expressed by words. I will

7 The definitions are "semantic" in the terminology of my survey of the varieties of definition (Davis 1986: §10.4).

8 Contrast Goldstein (1986), who argues that every term could be defined without specifying what kind of definition he has in mind.

argue in *Nondescriptive Meaning and Reference* that standard proper names are not definable in English or other natural languages, making it plausible that names express atomic ideas. Fodor generalizes this conclusion to the extreme, arguing that "all or most lexical concepts have no internal structure."[9] A "lexical" concept is one that is expressed by a semantically unstructured word in a given natural language. Thus the concepts "cat" and "brother" are lexical in English, in contrast to "feline animal" and "male sibling," which are phrasal. Fodor bases his lexical primitiveness thesis on the sort of evidence that Putnam marshaled against Katz concerning the definability of terms like "brother" and "cat," which we reviewed in §8.5. Fodor also points to the inability of generative semanticists to provide successful analytic definitions of terms like "kill," and the failure of epistemologists to define "know" despite a massive effort. It is well known that "To kill is to cause to die" and "Knowledge is true justified belief," for example, are too broad.[10] Attempts to strengthen the definitions tend either to produce circularity or to introduce material that is not known or intended by most people who use the terms "kill" and "know." Others point to the difficulty of defining everyday concepts like "cup" and "game."[11] While Fodor presumably intends his generalization to hold for all natural languages, and not just for English, it is nonetheless a contingent thesis about languages, with no fundamental implications about the nature of the mind. Fodor believes that the lexical primitiveness thesis has deep psychological significance, because he concludes from it that a large number of any normal human being's concepts are innate. But we will see in Chapter 17 that this conclusion depends on a false premise about concept acquisition.

9 Fodor 1981: 279, emphasis deleted. See also Putnam 1970a; 1973; 1975; Fodor 1975: 124–56; 1987: 161; 1994a: Chapters 2–3; 1998a: Chapters 3, 4, 5A; Fodor et al. 1980; Smith & Medin 1981: Chapter 3 ("The Classical View"); Lakoff 1987; E. E. Smith 1988: 21; Sterelny 1989: §5; Margolis 1998: §3; Millikan 1998b; Margolis & Laurence 1999: §2. Contrast Katz & Fodor 1963; Katz 1964b; 1972; 1974; 1977b; Schank & Abelson 1977: §1.4; Jackendoff 1989; Wierzbicka 1992a. For an introduction, see Lyons 1977: §9.9 and Chierchia & McConnell-Ginet 1990: 350–66.

10 Fodor (1998a: 71) also cites Quine's argument that there is no analytic–synthetic distinction given that no one has been able to provide a "serious and unquestion-begging" definition of the distinction. But this would prove much more than Fodor wants, since Quine's conclusion entails that even "unmarried males are unmarried" is not analytic, and that phrasal concepts are atomic too. Moreover, parallel reasoning should lead Fodor to the conclusion that "analytic" is simply a typical lexical concept.

11 See, e.g., Wittgenstein 1953: §§65–7; Smith & Medin 1981: 2–3; Lakoff 1987; E. E. Smith 1988: 21. Contrast Jackendoff 1989: 96.

In casual statements, Fodor sometimes says that *most* concepts are unstructured.[12] Without the "lexical" qualifier, this claim is unsustainable. Given the recursive procedures available for generating complex concepts, it is evident that there is an infinite number of complex concepts. We can count the number of lexical concepts, because the vocabulary or lexicon of any natural language is finite. But we cannot count the total number of concepts. Talk of percentages is therefore meaningless. The number of concepts that have ever occurred to humans is finite. Yet without a complete catalogue, or at least a completely catalogued sample, any estimate of the percentage of atomic concepts would be sheer guesswork.

Margolis and Laurence (1999) note that despite Fodor's advocacy, "conceptual atomism" is sometimes met with "stark incredulity." Part of the resistance, I suspect, is due to an equivocation about "concept." The term commonly means "conception," which denotes a type of belief system (see §19.4). Our conceptions of cats, brothers, and games are fantastically complex. As we can easily tell by introspection, it is part of our conception of cats that cats are four-legged, furry animals that meow and sleep a lot when adult. It surely makes no sense to say that a conception – that is, a belief system – has no parts. But we are using "concept" to mean "thought-part" rather than "conception." What the argument for Postulate 15.3 proves is that thoughts must have some cognitive parts that do not themselves have cognitive parts.

§15.3 THE CONCEPTION OF CONCEPTS

Definition 12.1 says that S is thinking T iff T is a thought occurring to S. The requirement that T be a thought is essential. For only propositions or other thoughts can be relational objects of thinking; yet thoughts are not the only event-types that occur to us. We have been observing that ideas other than thoughts – subpropositional concepts – also occur to us. So do nonrepresentational mental events such as pain, and nonmental events such as heart attacks. While we do not speak of thinking subpropositional ideas or concepts, we do speak of *conceiving* them. We conceive a concept or idea when it occurs to us. Thus it is natural to offer the following as a companion to Definition 12.1.

15.4 **Definition:** *S is conceiving C iff C is a concept (idea) occurring to S.*

12 E.g., Fodor 1981: 283; 1998a: 13ff. See also Margolis & Laurence 1999: 10.

Steve is conceiving the concept of the sky when the concept is occurring to him, as it must be if he is thinking the thought that the sky is blue. As with all of our definitions, we automatically extend Definition 15.4 in the natural ways to define other tenses of the verb "conceive." Thus Steve conceived the concept at midnight provided that the concept occurred to him at midnight.

Ockham and Reid held that a concept is the act of conceiving, not something separate at which the act is directed. Both adopted this theory in preference to the misguided Cartesian view that ideas are what we conceive *of* (see §19.1).[13] The Ockham–Reid view is almost correct on my conception of concepts. For on my view, concepts are event-types, and to conceive a concept is for it to occur to us. The concept of the Sun certainly cannot be identified with the Sun itself; we can conceive of the Sun, but it cannot occur to us. Ockham and Reid were wrong only on the subtlest of points: an event-type cannot be identified with the occurrence of that event-type, any more than a property can be identified with the possession of that property. If a concept were the act of conceiving, then conceiving the concept of a horse would be identical to conceiving the act of conceiving the concept, which is absurd.

Paralleling the definition of "thinking of" (Definition 12.2), we have:

15.5 **Definition:** *S is conceiving of Φ iff S is conceiving the concept of Φ.*

To conceive of the sky is to conceive the concept of the sky. Definition 15.5 holds even when the referent of "Φ" is itself a concept: to conceive of the concept of the sky is to conceive the concept of the concept of the sky. Note, though, that conceiving *the concept* of the sky is different from conceiving *of* the concept of the sky. Similarly, conceiving of the sky is an ordinary event (since we have the concept of the sky), whereas conceiving the sky is impossible (since the sky is not itself a concept).[14] The following forms, in other words, are not equivalent.

(1) S is conceiving Φ.
(2) S is conceiving of Φ.

When it occurs to me that the sky is blue, I am thinking about the sky but not about the concept of the sky. When it occurs to me that the

13 See Ockham, *Expositio*: 46; Boehner 1964: xxix; and Reid 1785: Essay 2, Chapter 4; Essay 4.

14 Nor an invention. The related sense in which Edison conceived the phonograph but not the sky will be discussed in §17.1.

concept of the sky is not blue, I am thinking about the concept. Form (2) differs from form (1) as much as "S is thinking of Φ" differs from "S is thinking Φ" (see §12.4). Definitions 15.5 and 12.2 yield as a theorem: *S is conceiving of Φ iff S is thinking of Φ*. Furthermore, Definitions 12.1, 13.1, 15.1, and 15.5 together entail that occurrent thinking is a specific kind of conceiving, since its object must be a proposition, or more generally, a thought. All thoughts are concepts, but not all concepts are thoughts.

H. H. Price argued that on a conceptualist view of thinking like mine, common thoughts would be impossible.

> [Y]ou have your occurrent idea of Dog . . . , and I have mine. They might be very similar, but they would still be numerically different. . . . It will be impossible to conceive the same occurrent abstract idea twice. . . . These consequences of the theory are rather disconcerting. We all want to say that many different people can 'think of the same thing'. . . . (Price 1953: 328–9)[15]

The concept of a dog, as we have said, is a *type of mental event*. Its tokens include the numerically different occurrences that Price referred to. You and I can both think of dogs because our numerically different mental events are tokens of the same type: they are both occurrences of the concept of a dog. Price's argument here is as fallacious as arguing that you and I cannot experience the same sensation – viz., pain – on the grounds that you experience your pain and I experience mine.

Like many other philosophers, Ryle found the existence of inner speech unproblematic, but denied that concepts could occur to us. He imagined that you spend five minutes finding the smallest prime number after 1300, and observed that "[Y]ou are working *with* the concept during the whole five-minute search, though the *phrase* 'prime number' comes off your lips, if at all, only once or twice during those five minutes"(1962: 448). This is evidence against the identification of thought with inner speech (see §19.3): thoughts and concepts are continuants, which can occur to us for some time, whereas sentences and words are discrete and episodic. Ryle, however, goes on immediately to draw the conclusion that there is no such thing as conceiving concepts!

> [T]he temptation to postulate the existence of special intellectual acts or experiences of 'concept-conceiving' or 'idea-having' needs to be resisted *ab initio*. . . . A stretch of thinking is not a procession of ephemeral incidents of having-concepts-in-mind. There are no such incidents, and if there were, a mere procession of them would not amount to thinking. (Ryle 1962:448)

15 Compare and contrast Geach 1957a: 13–14 and Crimmins 1992: §3.2.

This is completely unwarranted. Just because a concept occurs to us for a five-minute stretch does not imply that the occurrence is not an "incident," and indeed would seem to entail that we *can* conceive concepts, sometimes for as long as five minutes. The most Ryle could conclude is that your thinking of prime numbers, being a *continuous* process, cannot consist of a sequence of *discrete* occurrences. But even here, Ryle presents no evidence against the possibility that thinking seems to be continuous only because it consists of a sequence of discrete occurrences that are too close together in time to be distinguished introspectively.

Ryle's claim that thinking about things is not a procession of "having concepts in mind" is literally correct, although what he needed to show is that it does not consist of conceiving concepts. To have a concept in mind is not to conceive it but to conceive of it, which involves conceiving the concept of the concept. A related mistake was committed by what H. H. Price called the "classical theory" of concepts.

> Its essential tenet is that thinking is differentiated from other forms of cognition not only by being a special sort of activity, but also by having a special sort of *objects*, which are variously called universals, concepts, or abstract ideas. Thinking is conceived to be the inspection of such 'intelligible objects' and of the relations between them. (Price 1953: 301)

Literally, conceiving a concept or thinking a thought does not involve *inspecting* anything. To conceive the concept of a dog – that is, to think of a dog – is not visually or even introspectively to examine, view, or look at the concept of a dog. Examining a concept would involve being aware *of* it and thinking *about* it. Thinking about a dog does not entail thinking about a concept. The concept of a dog must be conceived, but it need not be present *to* the mind as an object of either perception or thought. Price himself made the reverse error: dubious (justly) about the existence of the extrasensory perception postulated by the classical theory, he inferred invalidly that "nothing which could be called a process of conceiving occurs" (Price 1953: 276, 316).

§15.4 CONTENT, OBJECT, AND REPRESENTATION

Concepts or ideas are said to *represent* things and to have *objects.* These terms can be defined as follows. Let "ϕ" stand for pronoun-free object nominals.

15.6 **Definition:** *i represents ϕ, and ϕ is the object of i, iff i = the idea (or concept) of ϕ, provided i is not a thought.*

Thus the idea of the sky represents the sky; its object is the sky, and its content is "the sky" (or *the sky*). It should be noted that "i represents *something*" is ambiguous. In one sense – the one just given – this expression means that i has a particular content: *i* is the idea expressed by the quantifier "something." In another sense, however, "i represents something" is tautological, and just means that i has some object: "*i* is the idea of ϕ" is true for some unspecified "ϕ."

If the sky is the object of an idea, then "the sky" is its content. That is, the idea whose object is the sky is the idea "the sky." As this example illustrates, the terms "content" and "object" are nearly coextensive in the case of the subpropositional ideas expressed by pronoun-free object nominals. Indeed, "'ϕ' is the content of i" and "ϕ is the object of i" are equivalent. We typically quote or italicize the term "Φ" when using it to describe the content of the idea, but not when using it to describe the object. I do not believe that any metaphysical inferences can be drawn from this difference in the usage of "content" and "object." It is just another manifestation of the fact that natural languages offer us a variety of ways to identify ideas using the words that express them (§7.6).

The terms "content" and "object" diverge significantly for other ideas, however. First, when Bill says "She gave me a kiss," the content of the idea that he used the pronoun "me" to express was "me"; but the object of the idea was not me (cf. the discussion of Definition 7.7). Second, the ideas expressed by syncategorematic terms have contents but not objects. The idea expressed by "is," for example, has the content "is." But the idea "is" is not the idea *of* anything. Third, "content" and "object" diverge substantively when applied to propositional ideas, but in a different way. In the sense cognate to Definition 15.6, the object of the thought that the sky is blue is the sky, whereas the content of the thought is "The sky is blue." In general, we can define content as follows.

15.7 **Definition:** *The content of i is* μ *iff i = the idea* μ.

This also holds when i is a nonpropositional thought. Since the thought "Will it rain?" is the idea "Will it rain?," its content is "Will it rain?" Nonpropositional thoughts are not said to have objects at all. It follows from Definition 7.9 that the content of i is "ϕ" iff the object of i is ϕ, where "ϕ" stands for any pronoun-free object nominal.

As is our custom with "the idea of ϕ," we will be interpreting "i represents ϕ" and its colleagues as *opaque* descriptions (§6.2). Consequently, they cannot be viewed as genuine relational predicates whose extensions contain all ordered pairs of the form ⟨i, ϕ⟩. The i position is fully referential,

but the ϕ context is intentional (see §6.3). Existential instantiation fails in the ϕ position: "The idea of Pegasus represents Pegasus" and "Pegasus is the object of the idea of Pegasus" are true, indeed logically true, even though Pegasus does not exist. Substitutivity of identity also fails: the thirty-fifth president is the object of the idea of the thirty-fifth president, but the thirty-fifth president is not the object of the idea of the president assassinated in Dallas, even though in fact the thirty-fifth president was the president assassinated in Dallas.

It is nevertheless common and natural to speak of the objects or contents of ideas as entities that ideas possess. It is certainly true that the idea of Pegasus has the content "Pegasus," and that the idea of the thirty-fifth president has a different content than the idea of the president assassinated in Dallas. Contents would appear to be mysterious fifth wheels, prime candidates for the metaphysician's razor. For we need to draw clear distinctions among the content "Earth," the word "Earth," and the planet Earth, allowing that the content "Earth" may differ from the content "the third planet" even though the planet Earth is the third planet, and allowing that the content "Earth" may be the same as the content "Erde" even though the words "Earth" and "Erde" are different. We also need to distinguish the content "Earth" from the concept or idea "Earth," and from the sense or meaning "Earth."[16] For one thing, concepts have contents, but there is no sense in which contents have concepts. For another, people have concepts but not contents. Finally, ideas can occur to people, with various causes and effects. But it makes no sense to say that a content occurs to a person. As for senses, words have senses but not contents, and ideas have contents but not senses. Senses are linguistic properties, contents are not. Contents are essential properties, senses are dependent on conventions.

If they are distinct from concepts, referents, words, and senses, what could contents be? The answer is simple: contents are *properties of ideas*. For an idea to have the content "Earth" is for it to have a property, a nonrelational property, and one that it alone possesses. The content "Earth" is the property of being the idea of Earth. The content of an idea is its identity, that which makes it the idea that it is. The content of an idea is thus markedly different from nonidentifying and nonessential properties, such as who possessed it, when it occurred to someone, which ideas it is associated with, and so on. It is precisely because contents are

16 Contrast Burge 1979b: 537; Peacocke 1986: 1–2, 5, 14, 48, 116; Fitch 1993: 471; and Boghossian 1998b: 257.

identifying properties of ideas that there is no theoretical need to refer to contents. Anything that can be said in terms of contents can be said directly in terms of concepts. The distinction between the idea "Earth" and the content "Earth" is no more important, or problematic, for psychological theory than the distinction between the Earth and the property of being the Earth is for astronomical theory.

The apparent contradiction between the claim that the content of an idea is a merely intentional object and the claim that the content of an idea is an identifying property of the idea can be resolved by noting that "the content of i" is ambiguous, with a referential and a nonreferential sense (cf. §21.1). In the nonreferential sense, the content of idea μ is μ. This is the sense defined by Definition 15.7. The content of the idea "Pegasus" is "Pegasus," which is not a real object in any sense. In the referential sense, the content of idea μ is the feature of *having the content* μ, which can be identified with the property of being the idea μ.

15.8 **Definition:** *i has the content* μ *iff* $i =$ *the idea* μ.

For the reasons noted earlier, "i has the content μ" cannot be interpreted as asserting a relation between i and μ, although it can be interpreted as asserting that i possesses an identifying feature. We can say similarly, it should be noted, that to have ϕ as its object is to be the idea of ϕ. Hence an idea has the content "the sky" iff it has the sky as its object.

As long as contents are understood in this way, we can say that *ideas are individuated by their content*. That is, i and i′ are the same concept if, and only if, they have the same intentional content. Ideas are similarly individuated by their objects, and by what they represent, as long as these notions too are opaque. This principle of individuation fails if ideas are defined as images, conceptions, or mental representations generally. There are many different images of Pegasus, for example, and there have been radically different conceptions of the Earth. It is also true, of course, that complex ideas are individuated by their constituents: i and i′ are the same only if they have the same components. But "content" in the sense related to "object" is not a mereological notion. Ideas *have* their contents and objects, but do not *contain* them.

§15.5 EXTENSION OR REFERENCE

Since we are interpreting "i represents ϕ" and "ϕ is the object of i" as opaque and intentional in the "ϕ" position, they do not express relations between i and ϕ (§6.3). We can, of course, consider the relation consisting

of the ordered pairs ⟨the idea of Aristotle, Aristotle⟩, ⟨the idea of Mars, Mars⟩, ⟨the idea of 2, 2⟩, and so on. I call this relation the *extension* or *reference* function, and refer to ex{i} as the extension or referent of i. The function needs to be defined differently for different kinds of concepts. We will focus only on singular concepts, those expressed by a singular term "σ."

15.9 **Definition:** *ex{i} = x iff for some σ, (i) i = the concept of σ and (ii) x = σ.*

Definition 15.9 says that the extension of an idea is determined by two factors: its content or identity, and the facts about what real objects are. It follows that the concept of the morning star and the concept of the evening star have the same extension, namely Venus (because Venus is both the morning star and the evening star). And the concept of the tenth planet has no extension (because there is no tenth planet). In a similar fashion, the extension of the concept of a dog would be the set of all dogs, and the extension of the concept of a thought would be its truth value.

We noted in §6.3 that many philosophers use psychological terms like "represent" relationally. The object or content of an idea, in this usage, is its extension or objective referent, if it has one. Taken relationally, ideas are *not* individuated by their contents or objects. For the idea of the morning star and the idea of the evening star are different but have the same extension, and the idea of the tenth planet exists but has no extension.

The ambiguous usage of "content" and its relatives creates difficulties when it goes unnoticed. Consider a recent conundrum in the philosophy of mind, which is how to resolve the contradiction in the following plausible claims.[17]

(3) Mental representations represent things.
(4) Representation is a relation between the representation and the thing represented.
(5) If representation is such a relationship, then misrepresentation is impossible.
(6) Misrepresentation is possible.

We are concerned with conceptual representation, so the mental representations we shall discuss are concepts. If we understand "represents" in

17 Cf. Cummins 1989: 27, 40, 57, 67; Fodor 1990a: Chapter 2; and Fodor & Lepore 1992: 157.

its conventional intentional sense, according to which "the concept of Φ represents Φ" is tautological, then (3) is trivially true. But in this sense, (4) is false. The concept of Pegasus represents Pegasus; but it is not true that the concept is related to Pegasus, since Pegasus does not exist. If we understand "represents" relationally, then premise (4) is trivially true. But in that sense, (3) is not universally true. It is not true in such a sense that the concept of Pegasus represents Pegasus. Only if we equivocate between the intentional and relational senses of "represent" does a problem arise (cf. Chapter 6).

Proposition (5) presents a different problem. If "represent" is interpreted relationally in the manner indicated, then no ideas represent nonexistent objects. However, ideas of nonexistent objects, such as the idea of Pegasus, are not usually described as "misrepresentations." Only propositional ideas can be misrepresentations. Before we can evaluate (5) for propositional representations, however, we have to decide what "things" propositions represent. In its most ordinary use, I believe, the concept that grass is green is said to represent *grass*. It represents grass in a certain way, namely, as being green. In general, *idea (or concept) i represents Φ as being Ψ if i = the idea that Φ is Ψ*. A more complex way in which an idea may represent grass as being green is illustrated by the thought that grass is green and pretty. This thought *contains and entails* the idea that grass is green, but is not identical to it. Precise conditions that are necessary as well as sufficient are hard to formulate for "i represents Φ as being Ψ." It appears that we may also say, for example, that the idea that grass is green represents greenness as being had by grass, or the grass's color as being green, or even grass's being green. Be that as it may, if we understand what propositional ideas represent in anything like this way, then (5) will fail even if representation is taken relationally. For (4) places no constraints on the way in which the thing is represented. It would allow that the idea that grass is purple (mis)represents grass as being purple.

To obtain a suitable relational sense of representation on which (5) is true for propositional representations, the things represented need to be entities like *facts* or *actual states of affairs.* So let us consider the relation consisting of ⟨the idea that the sky is blue, the sky's being blue⟩, ⟨the idea that the Earth is spheroidal, the Earth's being spheroidal⟩, and in general ⟨the idea that p, NOM(p)⟩, provided that NOM(p) is an actual state of affairs. If we call this the representation relation, then (4) and (5) are true. This is obvious, since *"representation" in this factive relational sense is just the correspondence relation*, which is the inverse of the satisfaction relation. But in this sense, (3) is not universally true. It is not true in this relational

sense, for example, that the idea that the Earth is flat represents any state of affairs. It does not represent the Earth's being flat, because the Earth is not flat. Moreover, (6) is trivially false. This observation does not imply that no one could possibly believe that the Earth is flat, or that those who hold such a belief are not mistaken, which would be completely absurd.[18] All that follows is that such beliefs cannot be said to represent or misrepresent any actual states of affairs, in the stipulated sense of "represent." Once again, only if we equivocate between intentional and relational senses of "represents" do we encounter a problem.

§15.6 THE CONTENT OF A THOUGHT

As already noted, the object of a thought in one sense is what we are thinking *of* or *about* when we think the thought. The content is *what* we are thinking, period. The content of the thought that p is simply *p,* or *that p,* or "p." Thus the object of the thought that the sky is blue is the sky, and the content is *(that) the sky is blue*. Alternatively, we can say that the content is "The sky is blue." The expression used to identify the content should be italicized or placed in quotation marks after "The content is," as in the previous sentences.[19] In general:

15.10 **Theorem:** *"(That) p" is the content of T iff T = the thought (that) p.*

When the thought is nonpropositional, "that" cannot be used. Thus the content of the thought "Win!" is "Win!"

In §12.1, we distinguished between "acts" and "objects" of thought – between the act of thinking and what is thought (or what is thought of). In §12.4, we distinguished between "intentional" and "relational" objects of thought. We said that the intentional object is what we are thinking *of*, whereas the relational object is what we are thinking. We emphasized that while the intentional object may not exist, the relational object must. We cannot think without thinking something. Expressions of the form "the content of S's thought" denote what S is thinking, the relational object of

18 There are arguments, from radically holistic premises, that misrepresentation is impossible because false belief and invalid inference are impossible (cf. Fodor & Lepore 1992: 160–1, 182–3). But these are not our concern here.

19 The content-clause is not usually italicized or quoted after "the thought." In "the thought *that the sky is blue*," the clause would be interpreted as an appositive, whereas it is interpreted as a subordinate clause in "the thought that the sky is blue."

S's thought. Hence when S is thinking "The sky is blue," the content of S's thought is the thought or proposition that the sky is blue. In general:

15.11 **Definition:** *T is the content of S's thought iff S is thinking T.*

A thought may be said to represent both its intentional object and its content. Clearly, thoughts are individuated by their contents, not necessarily by their intentional objects. The thought that the sky is blue has the same object as the thought that the sky is purple, but a different content.

Definition 15.11 can easily cause trouble. If S is thinking the thought that the sky is blue, this definition rules that the content of S's thought is the thought that the sky is blue. Since S's thought is the thought that the sky is blue, it appears to follow that the content of the thought that the sky is blue is the thought that the sky is blue. Definition 15.11 thus appears to produce the absurd result that the content of a thought is the thought itself, not merely an identifying property of the thought. Since Definition 15.10 rules that the content of the thought that the sky is blue is "that the sky is blue," we seem forced to take expressions of the form "that p" unnaturally to be singular terms denoting propositions.[20] We also seem forced to say that contents are events, which can occur to people and have causes and effects. These conundrums result, however, from a familiar fallacy of equivocation. In Definition 15.10, "thought" denotes an object of thought. In 15.11, "thought" denotes an act of thought. *The content of an act of thought is the relational object of that act, which is a proposition or other idea. The content of an object of thought is not the idea itself, but an identifying feature of the idea.*

The confusion is compounded by the frequent use of "content" in an objective, relational sense, to denote the state of affairs that would make a thought (or some other entity that objectifies the truth conditions of the thought) true.[21] The objective content of a thought is the state of affairs (or other objective entity) that the thought represents in the relational sense of "represents." I call this state of affairs the *situational extension* of the thought, in contrast to its *extension*, which would be its truth value. Thoughts are not individuated by their objective contents. The thought that the morning star is out and the thought that the evening star is out are different, but both would be made true by the same state of affairs, namely, Venus's being out. When "content" is used relationally without awareness of its intentional sense, trouble can result.

20 Cf. §13.3, fn. 13, and §21.1.
21 See, e.g., Stalnaker 1998: 340–2.

[I]t appears that [the propositional attitudes'] very nature as states that connect the organism to the world gives them a Janus-like character. On the one hand, in order to figure in the internal information processing, they seem to be necessarily individualistically construed; on the other hand in order to have content, they seem to be necessarily relational. Without content they seem to lack explanatory power, and without autonomous internal existence they seem psychologically and computationally inert. Therefore, some have suggested, perhaps the right conclusion from these conflicting considerations is that the concept of belief, and of propositional attitudes generally, is incoherent. . . . (Stillings 1995: 361–2)[22]

These contradictory demands can be resolved by noting that there is an intentional sense of "content" in which content is "individualistic" and not "relational." To say that the content of a belief or thought is *Santa is jolly* is not to say that the propositional attitude is "connected" to Santa or to his being jolly.

We will continue to give "content" and "representation" their intentional, nonrelational senses when we use them. More could be said to clarify these terms, but we will not be placing any theoretical weight on them. They are dispensable, and I believe their convenience for certain descriptive purposes is outweighed by the subtle distinctions they require and the confusions they tend to engender.

22 Cf. Johnson-Laird, Herrmann, & Chaffin 1984: 313. Contrast §6.2, this volume.

16

The Possession of Concepts

This chapter will define concept possession, and distinguish it from the stronger notions of understanding and mastery. To possess a concept, we will argue, is to have conceived it and remain capable of conceiving it. We will show that nominalist and information-semantic definitions of concept possession are wide of the mark, while recognition, knowledge, and inferentialist theories are too strong.

§16.1 POSSESSING CONCEPTS

Fodor has remarked that "[i]t's a general truth that if you know *what an X is*, then you also know *what it is to have an X*," and that this applies to concepts in particular (1998a: 2). There are many senses of "have" for which Fodor's generalization fails. A child may know what a baby is without yet knowing what it is to have a baby, and may know what a woman is without yet knowing what it is to have a woman. I believe that the having of concepts is a less obvious exception. I have defined concepts as parts of thoughts, but that does not tell us what it is to have a concept.

"Having" a belief is the same as believing the belief, and "having" a thought is the same as thinking it. To "have" an idea is to conceive it. Despite this pattern, *having a concept*, as this phrase is commonly used, must be distinguished from *conceiving a concept*. We have countless concepts that we are not currently conceiving. I have had the concept of neutrons since grade school. But I have conceived of them only on selected occasions since then, and was not conceiving of neutrons five minutes ago. Thus conceiving a concept is an occurrence, and concepts themselves are types of occurrences, even though having a concept is a "disposition." I believe that the latter may be defined as follows.

16.1 **Definition:** *S has concept C iff S has conceived C and is still capable of conceiving C.*

It follows that S has any concept that is occurring to him at the moment. For if it is occurring to him, then he has conceived it and is obviously capable of conceiving it. To have a concept that is not occurring at the moment is to have a previously activated ability to conceive it, that is, to have the persistent ability to reconceive it.[1] To have the concept of cats is to have the ability to think again of cats.

The sense of "ability" intended in Definition 16.1 is "*first potentiality*" rather than second – the sense in which someone who has learned to play the piano is able to play, while one who is merely able to learn to play is not. Our ability to acquire new concepts does not mean that we have them already. A pianist whose arms have fallen asleep is still able to play the piano in the intended sense, although in a stronger sense he is temporarily unable to. Similarly, someone who has repressed a concept is still capable of thinking it, although there is a sense in which he cannot. The second clause of Definition 16.1 implies that no matter how many concepts a man acquires during his lifetime, he has none after he dies. For he is then no longer capable of thinking. While death is the most obvious way to lose concepts, there are others, such as brain dysfunction. The process of retaining a concept is presumably related in some way to memory.

An immediate corollary of Definition 16.1, given Definition 12.2, is that *S has the concept of ϕ only if S has the ability to think of ϕ*. Another is that S acquired the concept of Φ only if S acquired the ability to think of Φ. It is natural to infer from this that the concept of Φ *is* the ability to think of Φ, and thus to adopt Evans's (1982: 104) view that concepts are abilities.[2] But concepts are constituents of thoughts, which occur to us. It would be a category mistake to maintain that abilities are constituents of thoughts, or that they occur to us. We cannot conceive an ability. Furthermore, having the ability to think of Φ is not sufficient for having the concept of Φ.[3] When "Φ" is an atomic concept, it seems evident that people have the ability to think of Φ only when the concept of Φ

1 This wording comes from Price's (1953: 323–4) formulation of the "classical theory" of concepts. In adopting this wording, we are not of course adopting the classical view that conceiving a concept involves *inspecting* anything (§15.3). See also Geach 1957a: 15; Evans 1982: 104; Cocchiarella 1984: 334–5; Fodor 1998a: 3; 1998b: 11.

2 See also Price 1953: 276–7; Geach 1957a: 12–15; Carruthers 1989: 106; Fitch 1993: 470–1; Millikan 1998b: 530–1, 537. Contrast Fodor 1998a: 125, fn. 6.

3 Contrast Fodor 1998a: 3; 1998b: 11; Moya 1998: 245.

has actually occurred to them. But when "Φ" is a complex concept, people have the ability to conceive the concept of Φ as long as they have all of its components and have conceived other concepts with the same structure. If S has the concept of Stradivarius violins and the concept of Steinway pianos, then he is certainly capable of thinking of Steinway violins.[4] But for most people, it would be an overstatement to say that they have the concept of a Steinway violin. Most people have never actually thought about Steinway violins – the idea has just never occurred to them (cf. §12.6).[5]

Both clauses of Definition 16.1 imply that Aristotle did not have the concept of a neutron. First, it never occurred to him: he never thought about neutrons. Second, he did not undergo the learning necessary to be able to think about neutrons. Aristotle not only *did* not, but also *could* not conceive of neutrons. In a psychological sense, something is *conceivable* provided that it is possible for the concept to occur to you. You must either have the concept or be able to acquire it. In a logical sense, something is conceivable iff it is possible for it to be true. The two senses are distinct: propositions can easily occur to us that cannot possibly be true, such as the proposition that $2 + 2 = 5$, or that no men are men. When philosophers such as Leibniz maintain that we cannot have an idea of an impossible thing like a square circle, they may have drawn this conclusion fallaciously from the undeniable premise that such things are inconceivable in the logical sense.[6]

Given the second law of occurrence (Postulate 14.2), according to which a thought occurs to S only if all constituent ideas occur to S, Definition 16.1 entails:

16.2 **Theorem:** *S is thinking P only if S has all concepts contained in P.*

Aristotle could not think the thought that neutrons are subatomic particles, because he did not have the concept of a neutron. The converse fails. I have had all of the concepts contained in "The moon is made of chocolate mousse" for some time. But to my knowledge I have never

4 Evans (1982: 103–4) dubbed this "The Generality Constraint." Cf. Fodor's (1987: 151) notion of the "systematicity" of thought, discussed in §14.2. See also Peacocke 1992: §2.1; Davies 1991: 239–40. Contrast Leibniz 1676: 281.

5 Philosophers should appreciate that no one had the concept "grue" before Goodman (1955: 74) introduced it, even though nearly everyone was capable of conceiving it: everyone with the concepts "green," "blue," etc. was able to think of an object that is green and examined before some future time or unexamined before then and blue.

6 See Leibniz 1684: 286–7 and Mates 1986: 68. Reid 1785: 429–30 draws the distinction clearly.

thought that thought before. The most we can say along these lines is that possession of all of the concepts in P gives us the *ability* to think P, provided that other things, such as excessive "length" or complexity, do not prevent it. Theorem 16.2 is concerned specifically with occurrent thought. The same principle holds for all propositional attitudes. Thus the first law of occurrence (Postulate 12.4) entails:

16.3 **Theorem:** *S believes P only if S has all the concepts contained in P.*

The two clauses of the first law of occurrence form the definition of concept possession. Since propositions are thoughts and thoughts are concepts, the first law can be reformulated as saying that S believes P only if S has P itself. Theorem 16.3 follows in virtue of the principle that having a complex concept entails having its components.

Definition 16.1 is not concerned with *conceptions* in the sense of §19.4. As we are using the terms, concepts are thought parts, whereas conceptions are belief systems. To have a particular conception, such as the Ptolemaic conception of Earth, is to have a particular set of beliefs. Having a conception of Φ is definitely sufficient for having the concept of Φ, by Theorem 16.3. For S has a belief about Φ only if S believes a proposition containing the concept of Φ. But it is not obvious that having a conception of Φ is necessary for having the concept of Φ. It is generally true, of course, that people with a concept of something have a number of beliefs about it. It is hard to imagine anyone with the concept of a cat who does not believe at least that cats are cats, and that any particular cat either is or is not black. But it is logically possible for people to have a psychological disorder that leaves them able to think thoughts without being able to have beliefs. It is also conceivable that coming to believe any proposition, no matter how obvious, requires at least some time during which the thought is processed in the way that results in belief or disbelief. But the only way that having a conception of Φ could be necessary to having the concept of Φ, given that thinking any proposition about Φ in any conception of Φ entails having the concept of Φ, would be if the propositions were believed just as soon as they were thought. Instantaneous belief may be possible, but it should not be a consequence of the definition of having a concept.[7] We will argue in §16.7 that having

7 For several reasons, Definition 16.1 is preferable to Bealer's definition, on which possessing a concept is equated with having a propositional attitude toward some proposition containing it (1998: 272). Having concepts is a *precondition*, not a mere logical consequence, of propositional attitudes.

a conception is necessary for the stronger relations of *understanding* and *mastering* a concept.

§16.2 NOMINALIST THEORIES

On my view, to have a concept is to have an ability, the once-activated ability to conceive the concept. Hence having a concept is a dispositional state, even though concepts are types of occurrences. Many attempts to define having a concept as a disposition or ability have failed because, under the influence of nominalism or behaviorism, the wrong ability was selected. According to one nominalist doctrine,

> To have a concept will at least be to understand the corresponding term. (Hamlyn 1967b: 140)[8]

As we saw in §13.1, and will stress again in §19.3, there is no essential connection between the concept of Φ and the word "Φ." A Russian may have the concept of a dog without being able to understand the word "dog." For he may not know a word of English. An adult speaker of English who suffers from aphasia may no longer be able to understand or use the word "dog," but that will not make him lose the concept of a dog or the ability to think of dogs. There is even good evidence that a child has the concept of a dog before she knows *any* words, and that a dog has the concept of food, or something like it, without being capable of understanding words. In addition to the anecdotal evidence for languageless thought, there is a large body of empirical evidence (see, e.g., Weiskrantz 1997). A further difficulty is that it is hard to explain what it is to understand a term, or to understand it in a particular way, except in terms of concepts. The fact that S uses the word "dog" extensively while using a code in which it means "KGB agent" is no evidence that S has the concept of a dog. What would show that S has the concept of a dog would be S's using (or understanding) the word "dog" *to mean "dog."* For S cannot mean "dog" by anything he said unless he was thinking of a dog. But we have explained meaning in terms of the expression of concepts, and have argued at length that concept is the more fundamental notion. Woodfield's (1997: 83–8) proposal that to have a concept is to

8 Cf. Berkeley 1710, 2nd. ed.: §27, 140, 142; Reid 1785: 193, 431, 532; Geach 1957a: 12–13, 16; Alston 1963b: 79; Aaron 1967: 196–7; Woodfield 1997: 94. Contrast Humphrey 1951: Chapter 8; Price 1953: 313–16, 344–6; Heath 1967: 179; Hamlyn 1971: 6; Vendler 1977: 58; and McGinn 1997: 102.

be committed to a communally recognized norm for classifying objects under a general term avoids some of these objections, but not all of them.

A nonbehavioristic form of nominalism was expressed more recently by Schiffer, for whom the dispositional state of belief is the underlying reality for concept talk.

> There are concepts, so to speak, but not as real entities over which one can quantify. To have the concept *dog* is just to have beliefs whose contents are correctly ascribable using 'dog'. (Schiffer 1987a: 270)

It is not clear why belief, rather than mere thought, or another propositional attitude such as desire, should be necessary for the possession of concepts. And even if it were true, Schiffer's analysis would not support the nominalist conclusion that he draws from it; he needs to show that the objects of belief are unstructured. But Schiffer's analysis is too weak: the belief that I did not have a hot dog for dinner tonight is correctly ascribable using the word "dog," but is no proof that I have the concept of a dog. Schiffer might try to avoid this by quantifying over senses. But is their reality any more secure than that of concepts? Indeed, we have defined senses in terms of concepts. Schiffer's analysis is also too strong. People had the concept "dog" before the English language arose. Moreover, the claim that *someone has a concept that is not expressed in English or any other natural language* is not self-contradictory. Such a claim is true, at least for a while, every time someone acquires a concept by observing a new kind of object.

§16.3 RECOGNITION THEORY

Another popular view is that to have the concept of something is to be able to *recognize* or *identify* it, or to *classify* things of that kind.[9] These three processes are very similar, but differ in some important ways. I take recognizing an F to involve perceiving an F and perceiving that it is an F. Perceiving that something is an F is one way of classifying or identifying

9 See, e.g., Price 1953: 35–6, 114, 260, 355; Geach 1957a: 36; Evans 1982: 270, 282; Johnson-Laird, Herrmann, & Chaffin 1984: 313; Carruthers 1989: 101, 106; Peacocke 1989c: 53–4; 1992: 110; Crimmins 1992: 94, 96; Millikan 1998a, 1998b; Cowie 1999: 139. Contrast Woodfield 1997: 86; Fodor 1998a: 104–5; 1998b. One of Fodor's (1998b) arguments is fallacious: If the concept of red were recognitional, that would imply that one could not have the concept of a red F without being able to recognize red things; but it will not follow that one can recognize red Fs (cf. Horwich 1998a; Grandy 1998; Horgan 1998). This is patent when "F" is something like "colorless apple."

it as an F, but classification and identification are more general processes that do not require perception. Physicists can classify particles that are too small to be perceived. The advice of an expert may enable Putnam to classify a tree as an elm, but more than such testimony would be needed to enable him to recognize elm trees. I will assume that any process of forming the belief that something is an F counts as classifying or identifying it as an F, although these terms may be used more narrowly. Recognition and identification are more general than classification in another way: the concept applied to the perceived object may be singular rather than general. I can recognize my wife when I see her as readily as I can recognize a woman when I see one. But the process is called identification rather than classification when singular concepts are applied. I assume that merely understanding an occurrence of the word "star," and recognizing that the speaker is talking about stars, does not count as recognizing, classifying, or identifying referents of the concept of a star.[10] Since recognition, identification, and classification involve the occurrence of concepts, these processes entail the ability to think and conceptualize. A man cannot recognize a wombat, for example, or classify something as a wombat, unless he can think about wombats.

What I will call the recognition theory, however, claims that the ability to recognize, identify, or classify referents is a necessary and/or sufficient condition of concept possession. Neither part of the thesis is plausible without qualification. The sufficiency claim requires only minor modification. The mere *ability* to recognize a Stradivarius piano is not sufficient for having the concept of one, although an actualized ability would be. Most attention has focused on the necessity claim.[11] No one requires the ability to recognize *every* referent of the concept under every condition of observation, only the ability to recognize at least *some* referents under some conditions. Even this claim seems to be false for a variety of concepts. First, as Geach (1957a: Chapter 7) observed, recognition theory fails utterly for logical concepts. The concepts expressed by terms like "and," "or," and "if-then" do not have referents that we can observe or recognize. The recognition theory also fails for what we might call "anti-recognitional" concepts. Consider the concept "unrecognizable virus,"

10 Contrast Millikan (1998b: 543), whose contrary assumption dilutes considerably the force of "recognition" theory she defends.

11 I am not concerned with the weaker claim that a recognitional ability is necessary for the concept to be *useful*. Contrast Millikan 1998b: 537. As truistic as this utility claim may seem, it is also debatable. Would it follow that the concept of God is "useless" if we assume that no one has the ability to recognize God?

defined as a virus that cannot be recognized (identified, or classified) as a virus. If science is as good as we hope it is, then there are no unrecognizable viruses. But whether there are any or not, we are now thinking about unrecognizable viruses, and therefore have the concept. By definition, though, it is not possible for anyone to recognize its referents.

Recognition theory is problematic even when we restrict ourselves to concepts that do have recognizable referents. The ability to recognize something depends on perceptual conditions that go well beyond possession of a concept. A woman who loses the ability to recognize red objects because she has gone blind does not thereby lose the concept of red. And since atoms and molecules are too small to be perceived, people cannot recognize individual water molecules in the same way that they can recognize glasses or pools of water. Heavy water can be recognized only through sophisticated tests. Recognition also depends on a variety of epistemic conditions, as do identification and classification generally. I am able to think about wombats, and I know some things about them: they are marsupials, they live in Australia, they are not cats or dogs, and so on. But I have never seen a wombat, or even a picture of a wombat, and do not know enough about wombats to be able to recognize one. A zookeeper out to trick me could easily lead me to classify a wombat as a coati, or vice versa. Recognition of most objects depends not just on having the concept, but on having a sufficiently well-developed *conception* of such objects (§19.4). It is also requisite that the subject lack undermining evidence or skeptical tendencies. A skeptic about the senses might refuse to believe that any object that he seems to see is really red, no matter how it looked or how much confirming evidence he had. The fact that he doubts that any object is red proves that he has the concept of red, even though his skepticism prevents him from classifying objects as red.

The recognition theory has its origins, I believe, in the classical empiricist doctrine that all concepts are acquired through abstraction or definition (see Chapter 17). According to the empiricist, the basic or atomic concepts are acquired through perception of their referents. Thus most people acquire the concept of red as a result of observing cherries, roses, and other red things. Now if any individual S did acquire the concept of F through abstraction, then it does seem true that S is, or at least was, able to recognize an F perceptually in the conditions that prevailed. However, S may retain the concept even after the recognitional ability is lost or blocked. Moreover, the classical empiricists overlooked some ways of acquiring concepts from experience. I acquired the concept of a wombat not through perceiving wombats, but through perceiving

statements about wombats. Any concept acquired through abstraction could have been acquired in this way through communication, and vice versa. So any recognitional ability implied by the process of abstraction does not attach to the possession of the concepts acquired by abstraction.

More fundamentally, even if recognition theory held for the possession of atomic concepts, the mere fact that other concepts are complexes of concepts for which the recognition theory holds would not tell us what it is to possess them (contrast Price 1953: 35, 358). Finally, concept empiricism is not an analytic truth. It is at least conceivable that some concepts are neither acquired by abstraction nor complexes of such concepts. The recognition theory cannot tell us what it would be to possess such concepts.

Some recognition theorists infer from their premise that having a concept is having a certain ability, to the conclusion that a concept itself is an ability. This conclusion has its own problems. As noted earlier, if concepts were abilities of any sort, they could not themselves be parts of thoughts. For thoughts are events. The view championed by Geach and Evans was that thinking consists of "*exercises*" of concepts.[12] To exercise the ability to Φ is to actually Φ. I exercise my ability to play the piano when, and only when, I play the piano. The identification of concepts with abilities would work fairly well if concepts were identified with the ability to have certain thought parts occur to us. For a thought occurs to us only if the ability of each one of its parts to occur to us is exercised. On Evans's view, though, concepts are *abilities to recognize or classify*. To exercise such an ability is to engage in some recognition or classification. But it is possible to think a thought without recognizing or classifying anything. When I think, "If Lassie had been a poodle, the television show would have flopped," I am not recognizing Lassie or classifying anything as a poodle. Hence I am not exercising my ability to recognize Lassie or to classify poodles. I am exercising my ability to think of Lassie and of poodles.

§16.4 INFORMATION SEMANTICS

Following Dretske (1981), Fodor sketched an "information semantics," according to which the content of a concept C is F iff occurrences of C covary with instantiations of F, and therefore carry information about

12 Geach 1957a: 12–14; Evans 1982: 101; Carruthers 1989: 106. Geach differed from Evans in claiming that concepts are abilities to use words.

F-ness. Hence to have the concept of F is to *have a mental state whose occurrence is reliably correlated with F-ness.* Misrepresentation can be allowed either by letting a reliable correlation be less than perfect, or by relativizing the correlation to standard or normal conditions. Fodor says that to acquire the concept of F is to get oneself "locked" to F-ness – to acquire a mental state that "resonates" to F-ness.[13] The basic idea is that conceptual representation is evidential meaning: to think of water is to token a natural sign of water.

In this unconstrained form, information semantics does not require that the content-bearing mental state be a concept occurring in the recognition process, and thus differs significantly from the recognition theory. But because of this generality, informational semantics is much too weak. Bill may experience fear when and only when he sees a grizzly. Then, in Bill, fear carries information about the presence of grizzlies. But that does not entail that Bill has the concept of a grizzly, and even if Bill's condition were universal would not give fear the content "grizzly." If anything covaries with instantiations of redness, the *sensation* of redness and the *belief* that something is red do. But neither of these distinct mental states is the concept of red. Information semantics could be restricted to saying that to have the concept of F is to have a *concept* whose occurrence is reliably correlated with F-ness. But then the theory would not tell us what it is to possess a concept. And it would seem to imply that a concept does not have its content essentially, since a state locked to F-ness could conceivably become locked to G-ness instead. But it makes no sense to say that the concept of red might become the concept of green.

Unconstrained information semantics is also too strong. While sensations of F are indications that a perceived object is F in standard conditions, and the belief that p possessed by a reliable authority is an indication that p, *mere thought* is not a natural sign of its objects. The fact that I am now thinking the thought that the tenth planet has a rectangular orbit is not a natural sign of any fact involving the tenth planet or rectangular orbits. It is mathematically impossible for my thinking about square circles to covary with the property of being a square circle. We would not expect

13 See Fodor 1990a; 1990b; 1991; 1994: 4–7, 57–60; 1998a: 12, 73–4, 125–6. See also Stampe 1979; Dretske 1981; 1995; Churchland & Churchland 1983; Millikan 1986; Maloney 1989: 134–9; Stillings et al. 1995: 341–2; Stalnaker 1998: 344–52; Margolis 1998; and Margolis & Laurence 1999: 60–1. Contrast Cummins 1989: Chapters 4–6; Sterelny 1990: 118ff.; Hatfield 1991; Goshke & Koppelberg 1991: 130–6; Loewer & Rey 1991: xxv–xxviii; Peacocke 1992: §5.2; Cowie 1999: 130–3; Margolis & Laurence 1999: 67–8; and Chapter 23, this volume.

the thought or even the belief that there are no tigers to covary with any fact about tigers.

The information semanticist seems to have focused implicitly on the occurrence of concepts in what we might call "*recognitional contexts*." Fodor could suggest that the content of a concept is determined by what it signifies naturally in recognitional contexts, while allowing that in other contexts its occurrence signifies nothing. In addition to importing the defects of the recognition theory, this hybrid has additional problems that are well known from the literature. The concept of water covaries with the property of being H_2O as strongly as it covaries with the property of being water. So the recognitional information theory rules incorrectly that (having) the concept of water is identical to (having) the concept of H_2O. Similarly, the concept of red covaries with redness to the extent it does because it covaries with the property of *appearing red*. The concept of a rabbit covaries just as strongly with the property of having undetached rabbit parts. The concept of an elm might covary more strongly with the disjunctive property of being an elm or a beech. Finally, the concept of Pegasus does not covary with the property of being Pegasus in any context. Fodor's optimism to the contrary, the basic facts that we have reviewed indicate that conceptual representation is fundamentally different from indicial meaning.[14]

§16.5 MEMORY AND KNOWLEDGE THEORIES

Price developed a sophisticated form of the recognition theory by emphasizing the role of memory in the possession of concepts.

> To have the concept of Cat is to have a (dispositional) memory of what cats are like. The basic way in which this memory shows itself is in the recognition of cats when I perceive them. (Price 1953: 355)

Concepts are dispositional memories, we could say, and concept occurrence is memory activation.[15] Price's theory is attractive because it emphasizes that the continued possession of a concept is a form of memory, and

14 The examples I have chosen respect Fodor's relativization to atomic concepts, and fail even his "asymmetric dependence" clause. Note that information semantics fares much better as a theory of *relational* rather than intentional content, since Russell's and Frege's problems no longer arise.

15 All of the elements were there, but Price did not make this last identification (see 1953: 329–30, 332, 356–7), probably because he focused on recognition memory rather than on recall memory. Recalling what a cat is like would surely imply having the concept of a cat.

uses the familiar distinction between standing and active memory to illuminate the distinction between the possession and occurrence of concepts. Price's theory does have difficulties, though. First, the initial occurrence of a concept cannot be the activation of a memory. Second, if the concept of a cat itself – and not just having that concept – is a memory of what a cat is like, then it is hard to see how the concept of a cat could be part of the thought that Tasha is a cat. A memory of what a cat is like does not have the form of the predicate of a proposition. Third, and perhaps most telling, it seems quite possible for a person to have the concept of a cat without remembering what cats are like, just as I have the concept of a wombat without remembering what they are like. In this case, I cannot "remember" because I never knew. But memory may simply fail. Someone may truly report, "I remember that when I was a child I saw a wombat in the zoo, but I cannot now remember what a wombat is like." Doing so would imply that he had the concept of a wombat, even though it does not satisfy Price's theory. Finally, Price himself observed that only basic empirical concepts can be treated as memory dispositions, without telling us what it is to possess other concepts, such as those of unperceivable objects (e.g., atoms, forces, perfectly straight geometrical lines), impossible objects (square circles), and pure abstractions (necessity).

Since remembering what Φ is like entails having a belief about Φ, it follows from Theorem 16.3 that *S has the concept of ϕ IF S remembers what ϕ is like*. So my analysis entails that Price's memory criterion is sufficient for having a concept. The main difficulties, we have seen, concern whether or not it is necessary for having a concept. A condition just as sufficient but more plausibly necessary is *knowing what (or who) ϕ is*. Whereas I do not remember or know what a Malipiero symphony is like, having never heard one or learned anything about them, I do know what a Malipiero symphony is. I knew this even when I first heard of Malipiero and his symphonies, and so could not yet be said to remember. The fact that the knowledge of what a Φ is is not predicatelike is still an obstacle to identifying this knowledge with the concept of a Φ, however. And it is unclear whether knowing what or who Φ is is necessary as well as sufficient for possession of the concept of Φ.[16]

16 Hamlyn (1971: 6, 16), Evans (1982: 74), Johnson-Laird, Hermann, & Chaffin (1984: 313), Davies (1991: 240–4), and Peacocke (1992: 22–4, 41–59) affirm the necessity, while Bach (1987a: 35, 40, 45) denies it. Vendler (1972: 132) does not say. See also: Mill (1869: 236), who equates conceiving something with *understanding* what it is; Smith & Medin (1981: 8), who say that "to have a concept of X is to know something about the properties of entities that belong to the class of X"; McDowell (1977: 170), who affirms only that some beliefs

It would be hard to find anyone with the concept of pain who does not know what pain is. It would also seem that knowing that a vixen is a female fox suffices for knowing what a vixen is. If so, then surely anyone who has the concept of a vixen knows what a vixen is. Even here, though, the inference would not seem to be logically valid. For surely the mere ability to think (which entails the possession of concepts) does not logically entail the existence of knowledge, and would not suffice to refute the skeptic. In other cases, furthermore, knowing who or what does not seem to be at all necessary for having a concept. For example, knowing that the president is the chief executive officer of the United States implies having the concept of the president; but it does not suffice in most contexts for knowing who the president is. And someone who has merely heard of Aristotle may have the concept of Aristotle; but at least among philosophers, that is not enough to know who Aristotle is. The criteria for knowing what or who Φ is are extremely relative.

> Whether a definite description . . . identifies an object for us rather depends upon our classificatory interests – on what we want to know. So, relative to an interest in locating him, we learn *who* the secretary of the bridge club is by learning that he is the person standing by the window. Relative to an interest in his name and occupation, we learn who he is by learning that he is Oswald Culbertson, author and ranking player. Relative to an interest in his performance, we may learn who he is by learning that he is the one who bid and made seven spades missing two aces. . . . (Loar 1976a: 363)[17]

Since whether or not someone counts as possessing the concept of the secretary of the bridge club does not vary in this way with our classificatory interests, it cannot be dependent on knowing who the secretary of the bridge club is. Reflecting the considerations summarized in this paragraph, the most extensive analyses of knowing who or what something is do not imply that such knowing is necessary for having the concept of that object.[18]

about Φ are necessary; and Burge (1979b: 537) and Jubien (1996), who assert that knowing what a Φ is or what it is for there to be a Φ to which we are related on different occasions is necessary for *mastering* the concept of Φ (see §16.7). Blackburn (1984: 94) observes that people cannot understand a term like "pain" without knowing what pain is, although they can understand "the richest man in the world" without knowing who the richest man is.

17 Cf. Linsky 1977: 91; Crimmins 1989: 287; 1992: 95; and especially Boër & Lycan 1975; 1986: Chapters 1–2.

18 See Chisholm 1976: 173–4 and especially Boër & Lycan 1975; 1986. Having the concept of N would seem to imply the "mental reference" conjunct of the preliminary Boër & Lycan analysis (at least when "N" is rigid), but not the knowledge or contextual importance

Gareth Evans proposed a related analysis that avoids the vagueness and relativity problems.

> Evidently a subject cannot be credited with such an Idea *a* unless he knows what it is for a proposition of the form ⌈δ = a⌉ to be true. So we can take the subject's Idea-of-the-object, *a* to consist in his knowledge of what it is for an arbitrary proposition of the form ⌈δ = a⌉ to be true. (Evans 1982: 110)

A special question about Evans's proposal is its implication that a subject cannot have any individual idea at all unless he has the concept of identity and truth. Are these concepts psychologically essential? To have the idea of *a*, would it not suffice for S to know what it is for *a* to be F, for any property F? A weaker form of Evans's view would require only that to have the concept of *a* is to know what it is for something *to be a*. But do people ignorant of atomic theory know what it is for something to be water? Do any of us know what it is for someone to be Aristotle?

§16.6 INFERENTIALIST THEORIES

Some have suggested that to have a concept is to have an *inferential ability*. Jonathan Evans said, for example, that "if I were not able to infer from the statement 'John is Mary's father, and Kate is the daughter of Mary' that John is the grandfather of Kate, then it would be hard to see how I could be said to possess the concept of 'grandfather'" (1989: 30). Peacocke (1992: 6) similarly maintained that to have the concept of conjunction is to find certain inferences (simplification and conjunction) primitively compelling because of their form. This approach is reminiscent of the "conceptual role theory,"[19] except that it focuses on the having of subpropositional concepts rather than on the content of propositional thoughts. The differences between Evans's and Peacocke's formulations will not affect the following conclusions.

Having the demonstrated ability to infer to or from any proposition containing the concept of Φ is clearly sufficient for having the concept

conjuncts (see Boër & Lycan 1986: 15). Similar observations can be made for their more refined analysis (Boër & Lycan 1986: 44, 89, 135).

19 See Sellars 1963; 1969; 1979: 78–83; Aune 1967b; Harman 1973: 43–6, 59–66; Field 1978: 13, 41–4; Loar 1981: §§4.2–4.6, 6.1, 7.3; Lycan 1981: 41; Schiffer 1982: 142–3; Stich 1983: 46–7; Peacocke: 1983: 106–11, 127–9; 1992: 6–8, 107, 131–51; Churchland & Churchland 1983: 304; Block 1986; 1993; Böer 1986: 103–10; Devitt & Sterelny 1987: 125–6; Taschek 1995a: 284–5; and Lance & O'Leary-Hawthorne 1997: 8–10. Contrast Stich 1983: Chapter 4; Peacocke 1986: 13; Sterelny 1990: §6.8; Fodor 1990a: Chapter 1; 1998a: 13–15; 35–7; and Fodor & Lepore 1992: Chapter 6.

of Φ. For inference is a process involving beliefs, and believing a proposition implies possession of all of its component concepts. But I do not believe that the ability to make inferences involving the concept of Φ is necessary for possession of the concept. Evans is certainly correct that in ordinary circumstances, possession of a concept implies the ability to make appropriate inferences. But it does not seem to be either logically or psychologically impossible for a subject to suffer from a psychological block or neurophysiological disorder that leaves him temporarily unable to make inferences in a particular conceptual domain. During such a period subjects might report, for example, that they can think of various questions about grandfathers, but cannot form an opinion about any of them, even though the answers may seem tautological to us. After the problem goes away, subjects might be amazed at how easy the questions were. Such a possibility cannot be ruled out on the basis of any empirical evidence that I know of, and is not the sort of possibility that can be ruled out a priori.[20] Evans justified the connection that he claimed between concepts and inferential abilities by linking the latter to *understanding*. This link is plausible, but as we will see in §16.7, understanding or grasping a concept is stronger than mere possession.

Even if possession of the ability to make certain inferences involving the concept of Φ were both necessary and sufficient for possessing the concept, possessing the inferential abilities would not be *what it is* to possess the concept of Φ. For the ability to make inferences involving the concept of Φ is a *consequence* of possessing the concept of Φ. I have the ability to infer "Some passed" from "Everyone passed" only because I have the ability to think and believe these two thoughts. I have the ability to think and believe those thoughts only because I have their constituent concepts.

§16.7 UNDERSTANDING AND MASTERY

The *possession* of concepts is an all-or-nothing affair. We cannot speak of the degree to which a concept is possessed, nor of strengthening or weakening our possession of it. We cannot strengthen our possession of a concept once we have acquired it. Our *grasp* or *understanding* of concepts,

20 Peacocke's focus on *primitive* compulsion is also questionable. Imagine a person who always inferred "p" from "p&q" as follows: p&q; ∴ "p" is true; ∴ p. Or as follows: p&q; ∴ – – p; ∴ p. Such a person must have the concept of conjunction to make these perfectly reasonable inferences.

by contrast, does vary in degree.[21] While I possess both the concept of a neutron and the concept of a quark, my grasp of the former concept is much greater than my grasp of the latter. My grasp of the concept of relativity has slipped since I took physics in college, while my understanding of the concept of probability has strengthened. I am presenting a great many facts about ideas in Part III so that my readers will have a firm enough grasp of the concept of an idea to understand and evaluate the theory of meaning and expression that I present.

The *extent* to which S grasps the concept of Φ is determined by the depth, breadth, and accuracy of S's conception of Φ. That is, S's grasp of the concept of Φ depends on how much S knows about Φ, and on how fundamental what S knows about Φ is. I believe that the simple, qualitative claim that S *grasps* or *understands* the concept of Φ is true provided that S grasps the concept to a *sufficient* extent, that is, provided that S's knowledge of Φ is sufficiently broad, deep, and accurate. A concept is not itself knowledge, but knowing a proposition containing that concept puts us in a particular relation to it. Having sufficient knowledge centered on the concept puts us in the relation called understanding the concept.

We observed earlier that "concept" is ambiguous in English, and often means a centered belief system rather than a mere thought-part. I call such systems "conceptions." Our conception of Φ consists of our beliefs about Φ (§19.4). In these terms, *to grasp or understand the concept of ϕ is to have a sufficiently extensive and accurate conception of ϕ*. It follows from the definition that *grasping a concept implies possessing it*, since having knowledge or beliefs about Φ does. I believe that the converse fails, although the issue is unclear.[22] If someone knows that quarks are particles discovered by physicists because he went to school, then he has the concept of a quark. He can think about quarks, and thus conceive of them. But if that is all that he knows about quarks, then I do not believe that he qualifies as grasping or understanding the concept of a quark. I barely understand the concept, yet I at least know that quarks are the constituents of protons, electrons, and neutrons. Furthermore, if possessing a concept entailed grasping it, that would imply that acquisition of the concept of Φ could not have preceded acquisition of knowledge about Φ. This would

21 Cf. Plantinga 1985: 356; Maloney 1989: 56–7; Higginbotham 1998.

22 What is especially controversial (and groundless in my opinion) is the holistic thesis that possessing the concept of Φ requires having a *substantial*, and predominantly *accurate*, conception of Φ. See Fodor & Lepore 1992 for discussion and references. For authors who assume that possessing a concept implies grasping it, see Hamlyn 1971 and Gillett 1992: Chapter 1.

mean that the subject could know that Φ is F, say, without having had any time to consider whether Φ is F, make observations, weigh the evidence, or engage in rational reflection. This would only seem possible if the knowledge were truly self-evident. But the knowledge required for understanding a concept often seems far from self-evident.

How much knowledge about Φ is necessary for someone to understand the concept? I am not sure that a threshold of sufficiency can be specified, and I suspect that it has the same context relativity that the notion of knowing what Φ is has (§16.5). Thus before I would judge that he understood the concept, a graduate student would have to display more knowledge of functionalism on an exam than an undergraduate would have to display. Similarly, I believe that in a legal context, the knowledge that marriage involves certain legal arrangements would be essential to counting someone as understanding the concept of marriage, but not the knowledge that marriage is a Catholic sacrament. In religious contexts, the latter knowledge would be essential.

The thesis that grasping the concept of Φ consists of having a sufficiently extensive and accurate conception of Φ is not incompatible with the idea that it requires certain recognitional capacities (Price 1953), inferential abilities (J. Evans 1989), or evidence-gathering techniques (Gillett 1992). For having a sufficiently extensive and accurate conception plausibly requires such abilities.

To *master* a concept is to *fully* understand it, which requires an even higher level of knowledge than understanding or grasping *simpliciter*. I only barely grasp the concept of a quark – I am far from fully understanding it. Putnam, as is well known, grasps the concept of an elm, but he is far from mastering it, since he does not know the difference between an elm and a beech. Again, I believe the level of knowledge required for full understanding is context-dependent, and doubt that it can be precisely specified for any context. It is patent that concept possession does not entail concept mastery. Mastery of the concept of relativity, for example, requires a lengthy period of learning, during which one is thinking about relativity but does not yet know enough about it.[23] It is also evident that concept mastery does not entail omniscience. Most people master the concept of a prime number early on, but no one – not even the most

23 Higginbotham's (1998) work raises the question of whether mastery of the concept of Φ requires an adequate conception of the concept as well as of Φ. I have seen no evidence to suggest that mastery of the concept of elm trees requires knowing anything about concepts rather than trees.

accomplished mathematician – ever learns everything there is to know about prime numbers.

Peacocke's work illustrates the importance of distinguishing possessing from mastering. Peacocke seeks to identify what makes it true that a subject has one set of thoughts or propositional attitudes rather than another. Since he defines concepts as constituents of thoughts ("contents"), this leads him to search for the conditions under which a subject possesses one concept rather than another. He proposes, for example, that

> The concept *red* is that concept *C* to possess which a thinker must meet these conditions: 1. He must be disposed to believe a content that consists of a singular perceptual-demonstrative mode of presentation *m* in predicational combination with *C* when the perceptual experience that makes *m* available presents its object in a red′ region of the subject's visual field and does so in conditions he takes to be normal, and when in addition he takes his perceptual mechanisms to be working properly. . . . (Peacocke 1992: 7)

Peacocke observes that it might well be true that a subject believes of a genuinely red object that it is not red, because the subject has slightly misidentified the range of colors that are red. The subject believes that apples, stop signs, blood, and the like are all red, but for some reason he believes that cherries are not red. The congenitally blind have lots of (correct) beliefs about red objects. Yet if they have just gained sight, an apple may be presented in a red′ region of their visual field without their being able to recognize it as red; they may mistakenly conclude that the apple is not red.[24] Since these subjects have attitudes toward propositions containing the concept *red*, they must possess that concept. But the subjects do not satisfy Peacocke's "possession condition." Peacocke defines a possession condition at this point as one that "states what is required for full mastery of a particular concept" (1992: 29). Since the subjects we are considering have not fully mastered the concept *red*, Peacocke's condition seems plausible. This leads him to conclude further that a thinker may have attitudes toward thoughts containing the concept *red* without fulfilling its possession condition. But it is absurd to claim that someone can think that something is or is not red without possessing the concept *red*. Even Peacocke himself goes on to say, for example, that "[s]ince the thinker is capable of entertaining the thought *Fa*, he possesses the concept *F*" (1992: 43). This contradiction can be resolved by observing that

24 Note that the vast majority of color concepts do not satisfy Peacocke's recognition condition even with respect to people with normal sight, as anyone who has spent time with color charts in paint stores can attest.

possessing a concept is different from mastering or fully understanding it. Thinking a thought containing a concept entails possessing the concept, but does not entail fully understanding it. Peacocke's conditions are plausible for mastering a concept but too strong for merely possessing it.

The distinction between grasping and merely possessing a concept similarly helps us to see the element of truth in Crimmins's account of having a concept.

> 'Having a concept', I suggest, is simply having a *normal* idea. Typically, this means that one needs, among other things, normal beliefs and recognition abilities to count as having the concept. Knowing about the potent solvent 'H_2O' is not sufficient for having the concept of water. One must know that water is a clear, more or less tasteless, colorless liquid at room temperature, that it is safe to drink, and so on. (Crimmins 1992: 94; see also 1989: 287)

But it is quite possible for normal people with exposure only to polluted water, not to mention cranks exposed to pure water, to believe that water is poisonous and brown. Since anyone who believes that water is poisonous has to have the concept of water, *having* a concept of water cannot entail having a normal conception of water. But since *grasping* or *mastering* the concept entails having a sufficiently extensive and accurate set of beliefs about its object, we can surely judge that those who do not have what Crimmins calls a normal idea of water have not mastered the concept. If their conception of water is sufficiently deficient, it will even be true that they have a poor grasp of the concept of water. It should be noted that "normal" cannot be interpreted purely statistically here. It is quite possible that most of what statistically normal people believe about a subject is false, in which case they have a poor grasp of the concept.

17

The Acquisition of Concepts

Chapter 15 explored the basic properties of ideas or concepts, and Chapter 16 defined what it is to possess them. This chapter looks at how we come to possess concepts. We will identify four sources: observation (abstraction), communication, constructive thought, and abstractive thought. Psychological investigation of concept formation has focused on a process called "concept learning." Since the process involves the formation and testing of hypotheses containing the concept to be acquired, it cannot be a way of acquiring concepts. This fact has been used to argue that all concepts are innate. But the argument assumes groundlessly that concept learning is the only method by which concepts could possibly be acquired. The proper conclusion to draw is that since concepts are thought-parts rather than beliefs, the acquisition of concepts is not a belief-forming operation. We will explore the possibility that some concept-formation processes, involving observation and abstraction, are basic psychophysical processes. I believe that the concentration on concept-learning experiments can be traced back to the historical confusion between the acquisition of concepts and the learning of word meanings.

§17.1 ACQUISITION

We are said to *acquire* or *form* a concept when we change from nonpossession to possession.

17.1 **Definition:** *S acquired (formed) concept C at t iff S possessed C at t but not before.*

Acquisition coincides with an initial occurrence marking the beginning of a period of possession, during which the concept may or may not reoccur.

Thus "forming a concept" (which implies its initial occurrence) has a more specific meaning than "conceiving a concept" (its occurrence or reoccurrence). Although it is logically possible to lose a concept and then reacquire it, we normally form a concept just once, when it first occurs to us. In the normal case, a concept acquired at t has never been possessed before t. In every case, there must be some period of nonpossession ending at t.

Several processes result in the formation of concepts: observation or abstraction, communication, constructive thought, and abstractive thought.

(1) *Observation*. We may acquire the concept of a platypus simply by seeing platypuses in the zoo, or by seeing pictures of them in a book. We may acquire the concept of an afterimage by being aware of one introspectively. Observing an object often causes us to think about it even though we have never thought about it before. This process is traditionally called *abstraction*,[1] a usage we will follow without thereby endorsing traditional theories about how abstraction works. Note well that as long as a woman is not blind, she can *see a platypus* before she has the concept of a platypus, although she cannot beforehand see *that* something is a platypus or see something *as* a platypus.[2]

(2) *Communication*. We may acquire the concept of a wombat by hearing other people talk about wombats, or by reading about wombats. This may occur even though we do not pick up enough information to visualize or define wombats. Communication was presumably ignored in classical discussions of concept formation because it is not an original source of concepts among human beings generally. Before a concept can be communicated to others, it has to be acquired by someone through another process. But communication is responsible for a large portion of any individual's stock of concepts. The fact that communication dramatically expands the range of our thoughts is one of the fundamental advantages of social living.

(3) *Constructive Thought*. We may acquire the concept of a winged earthworm by simply imagining one – by mentally "putting together" the concepts of wings and earthworms. The imagination may be sensory, or purely conceptual. As an example of the latter, I might mentally combine the

1 Cf. Aristotle, *De Anima*: 431b15; Aquinas, *Commentary on De Anima*: Lns III.10.54–63; *Treatise on Man*: Art. I.79.3; Locke 1690: §§3.3.6–11; Reid 1785: Chapter 5.3; Mill 1879: Chapter 4.2; James 1890: Chapter 12; Joseph 1916: 34–5; Humphrey 1951: 265–78; Price 1953: 38–9, 215; Rand 1969: 15. Contrast Geach 1957a: Chapters 6–11; Carruthers 1992: 54ff.

2 See Dretske 1969. Contrast Sellars 1963: 176; Carruthers 1992: 55; Fodor 1998: 136, fn. 10.

concept of a prime number with that of a number less than 10 to conceive of a prime number less than 10. I might coin the term "first-order prime number" to denote such a number. If I continue to define "second-," "third-," and "fourth-order" prime numbers in terms of 100, 1,000, 10,000, and so on, I might then wonder whether the number of prime numbers increases or decreases with their order. The constructive thought resulting in possession of a new concept may itself result from communication. If a speaker expresses a complex concept that the hearer did not possess, the hearer may acquire the concept when she grasps what the speaker meant. The process of combining concepts to get new concepts is often described as "definition." But definition involves either the assignment of concepts to words, or the expression of concepts in words. Hence there are two key differences between definition and constructive thought. The defined concept must exist in the definer's mind before the act of definition. And constructive thought need not involve words.

(4) *Abstractive Thought*. Once some concepts have been acquired, it is possible to acquire others by abstraction without the use of observation. Thus a person could acquire the concept of a raptor by thinking of hawks, falcons, owls, and ospreys, and contrasting them with robins, sparrows, pigeons, and ostriches, without at the time seeing, hearing, or otherwise observing any birds. In this case, the subject would acquire the general concept of a raptor as a result of thinking of particular raptors. Occurrences of particular concepts cause formation of a more general concept. It is possible that abstractive thought is simply a special case of constructive thought. This would turn out to be true, for example, if thinking of hawks, falcons, and owls caused the subject to focus on their talons, hooked beaks, and carnivorous habit, and if the concept of a raptor were analyzable as the concept of a carnivorous bird with talons and a hooked beak. But as the arguments of Putnam, Fodor, and others have made clear, it is doubtful that the concept is definable in this way (see §15.2). Whereas constructive thought invariably produces *complex* concepts, abstraction generally produces *atomic* concepts, and communication often does.

A special case of concept formation through constructive thought is *concept creation*. To create the concept of Φ is not just to think of Φ, but to *think Φ up.* When S has created the concept of Φ, the linguistic forms *S conceived Φ* and *S conceived of Φ* are both true in a sense stronger than the one defined in §15.3. In this more specific sense, Thomas Edison conceived the phonograph, but not the sky or the automobile. This is true even though there is a standard sense in which Edison did conceive the concept of the sky. He thought of the sky often, surely, even though

he did not think up the sky. In the creative sense, to conceive the light bulb is to conceive of the lightbulb: *S conceived Φ* and *S conceived of Φ* are equivalent. In the general occurrence sense, it makes no sense to speak of conceiving a lightbulb; only concepts can be conceived. "S had an idea" also has a strong sense in which it denotes not just the occurrence of the idea, but its novelty. Note that to conceive something in the creative sense is not to invent it. Leonardo da Vinci conceived the helicopter, but he never got to create a working model. The patent office is filled with conceptions that never saw the light of day.

Recall that concept possession requires two things: the possession of an ability, and its exercise. That is, a concept is possessed only if the concept is able to occur to the subject, and has occurred to the subject (Definition 16.1). The ability required is a first rather than a second potentiality. We must have the ability to think of the concept's object in the sense in which we do not have the ability to play the piano until we learn to play, although we had the ability to learn all along. Since concept possession requires both an ability and its exercise, there are two ways to acquire a concept, depending on whether the ability is acquired before it is first exercised, or at the same time. Concepts may be acquired when the underlying ability is acquired, or afterward. In the case of atomic concepts produced by abstraction or communication, the ability to think of the novel object appears to be acquired at the same time that the concept is acquired. Thus when people acquire the concept of a platypus by observing one, they do not appear to have the ability (first potentiality) to think of a platypus before they actually begin to think of one. The ability seems to be actualized as soon as it is acquired. In the case of concepts formed through constructive thought, by contrast, the ability to think of the object may be acquired long before the concept is acquired. Thus as soon as a man acquires the concept of a wombat, he will normally acquire the ability to think of a blue, rabid, winged wombat. But he may not actually think of such a thing for a long time, if ever. He cannot be said to have formed the concept of a blue, rabid, winged wombat until he actually conceives that concept. There appears to be a *third law of occurrence* (§12.6, §14.1), therefore, that applies specifically to atomic concepts: *S is able to conceive an atomic concept only if the concept has occurred to S.* S will have the ability to conceive a complex concept, however, as long as S has all of the simple constituents, and has conceived other complex concepts with the same structure. For an analogy, we do not credit anyone with the ability to play a keyboard instrument until they have actually played one. But we do credit musicians with the ability to sight-read particular scores

that they have never sight-read before. Conceptual abilities, like musical abilities, are productive.

§17.2 CONCEPT LEARNING AND INNATE IDEAS

Empiricists maintain that abstraction is the only original source of atomic concepts, and that all concepts are therefore *empirical*. The scholastic maxim was "Nothing is in the intellect that was not first in the senses." Rationalists maintain that there are nonempirical sources, which produce a priori concepts.[3] Common examples of putative a priori concepts are necessity, God, and logical concepts. We have no stake in this issue.

Rationalists and empiricists have engaged in a lively debate as to whether there are any *innate* ideas.[4] It is agreed by all parties that we have an innate ability to form concepts under certain conditions. The dispute concerns whether any ideas themselves are innate. The controversy has been clouded by the existence of different definitions of innateness. (1) Locke took "innate" to mean "possessed from birth." On this definition, children would have to be born thinking of the objects of innate concepts, making it evident that few if any concepts are innate in this sense. While Locke was reacting to influential contemporaries, few if any major philosophers have maintained that concepts are possessed from birth. So this definition is of limited contemporary interest. (2) A more general definition, suggested by Leibniz's formulation, takes a concept to be innate as long as the ability to conceive it is possessed at birth. On this understanding, an innate concept may still have to be acquired. Acquisition would occur when the latent ability is first manifested, that is, when the concept is first conceived. Experience may be necessary to "trigger" the first occurrence of the concept, even if the ability to conceive it is there all along. Concept acquisition would then be more like remembering than learning, although neither term would apply. It is not nearly as implausible to suggest that newborn children have the unactualized ability to think of numbers and other mathematical objects as it is to maintain that they actually think of them. So nativism is easier to defend on the Leibnizian definition than on the Lockean. Still, it is hard to believe that babies already have the ability to think of prime numbers,

3 See Adams 1975: 78–86; Burks 1977: 608–19.

4 Descartes 1647: 448; Locke 1690: Book I; Leibniz 1709; Yolton 1956: Chapter 2; Hacking 1975: 57–69; Stich 1975; Fodor 1975: Chapter 1; 1981: Chapter 10; 1998a: Chapters 6 and 7; Samet & Flanagan 1989; Carruthers 1992: 50; Sober 1998; Cowie 1999.

or that they are born with the ability to think of telephones. I do not believe that a definitive answer will be available until we know the neurochemical basis of the distinction between thinking of something, having the ability to think of it, and having the ability to acquire that ability. But innate characteristics presumably must be inherited, and thus would have to be products of natural selection. Since the cognitive abilities in question would have conferred no survival or reproductive advantage on the human species when it evolved a million-odd years ago, it is hard to see how they could have been naturally selected.[5] (3) Descartes's formulation suggests an even more general definition, on which a concept is innate if the subject is born with a disposition to acquire it under certain conditions. On this definition, it is uncontroversial that most, if not all, concepts are innate. But the term has been emptied of its customary content. Adams (1975: 74–8, 84) observes that Descartes's theory was nevertheless significant in its denial of the Aristotelian theory that ideas are formed when the forms of external objects literally enter the mind through perception. Adams goes on to observe that Descartes also rejected Locke's doctrine that all ideas are empirical, and suggests that Descartes often uses "innate" in a narrower sense to denote a priori ideas that we have a natural ability to form. (4) Recent authors inspired by Chomsky and biology have strengthened Descartes's formulation by restricting the term "innate" to concepts that all normal humans are disposed to acquire in the normal course of development. Many concepts would appear to be innate on this definition (e.g., basic color, numerical, linguistic, and psychological concepts), but not all (e.g., modern scientific and technological concepts). A definitive classification would depend on a clear idea of what is considered normal. But it should be true on any reasonable interpretation that while many concepts are normally acquired, some are not.

The issue of whether there are innate concepts in some strong sense does not bear on the central questions we are concerned with. But it will be instructive to consider Fodor's "standard argument" that concepts must be innate because they cannot be "learned."[6] "*Concept learning*" has a well-entrenched technical meaning in the psychological literature.[7]

5 Sterelny 1989: 122; 1990: 130; Samet & Flanagan 1989: 205.

6 See Fodor 1975: 34–42, 89–97; 1981: 266–9; 1998a: Chapters 6 and 7; Fodor et al. 1980: 504–5. See also Geach 1957a: 42–4; Carruthers 1992: 55; Cowie 1999: 72. Contrast Sterelny 1989: 130–2; Maloney 1989: 102–3, 120–1; Samet 1986; Samet & Flanagan 1989.

7 See Hull 1920: 5–6; 25–31; Humphrey 1951: 272–6; Bruner, Goodnow, & Austin 1956: 22–3; Hunt 1962, esp. Chapter 1; Benjafield 1992: 64.

In a typical experiment, the subject must learn to produce a designated response under set environmental conditions on the basis of reinforcement or feedback. The subject may learn to say "Flurg" when and only when a green or square object is presented, after being informed that his guesses were correct or incorrect in a series of trials in which objects of various colors and shapes are presented. A typical finding is that mastery of this task takes longer when the criterion is disjunctive than when it is conjunctive. The classical theory is that the subject formulates tentative hypotheses about what the concept "flurg" is, and revises them until every response is correct.

While *something* is certainly learned in such experiments, the mechanism proposed cannot account for the *acquisition* of *concepts* (i.e., thought-parts). The subject may, for example, have acquired the following beliefs:

A flurg is a green or square object.
The term "flurg" applies to an object iff it is green or square.
"Flurg" means "green or square object."
The concept "flurg" is the concept of a green or square object.
The concept the experimenter has in mind is that of a green or square object.

But the subject could not have entertained any of these propositions, even as tentative hypotheses, without possessing the concept of a flurg – that is, the concept of a green-or-square object. The problem is particularly clear in the case of atomic concepts: acquisition of the concept *green* cannot be the result of forming hypotheses about green objects, for the hypothesis cannot be formed unless the concept of green is already possessed. In general, since concept learning is defined as a process involving the formation of *hypotheses about things falling under a concept,* it cannot be used to acquire *that* concept.[8] That concept must be possessed before any hypotheses containing it are entertained. This is a consequence of the laws of occurrence, specifically, Theorem 16.2.

The same laws imply that the process of acquiring a *conception* (belief system) by the process of hypothesis testing and confirmation cannot account for the acquisition of any of the *concepts* (thought-parts) involved

8 It is possible to formulate hypotheses about a concept without possessing that concept if the hypotheses contain second-order concepts. But formulating such a hypothesis will not generally result in acquiring the first-order concept that it is about. Thus I might formulate the hypothesis that an object falls under *the concept John expressed using the nonce word "bliff"* iff it is the square root of 49. The concept I must have to formulate that hypothesis is the *concept of the concept* John expressed using "bliff." I need not have the concept John expressed using "bliff," and formulating that hypothesis will not result in my acquiring the concept John expressed if it is, say, the concept of the third prime number.

in that conception (§19.4). This holds true whether the conception has the structure of a "definition" or a "prototype" structure. Acquiring the concept of a dog cannot require learning that collies are typical dogs, or that dogs are typically four-legged, as Sterelny (1989) and Cowie (1999: 77–8) suggested. For learning what is typical of dogs requires possession of the concept of a dog. More generally, acquiring a concept cannot depend on adopting any propositional attitude whose relational object contains that concept. Thus Fodor's (1994: 35) suggestion that you can acquire the concept of an elm by "making it a policy" to think of an elm iff an expert does, suffers from the same defect. The indicated propositional attitudes presuppose that the concept in question has already been acquired.

Fodor appears to believe that the standard argument applies only to atomic concepts, and that complex concepts can be learned "inductively."[9] But having the separate concepts is not enough to form the requisite hypothesis: they must be put together in the right way, as the "flurg" example shows. Any hypothesis in which the component concepts are put together in the right way contains the complex concept. So the laws of occurrence again entail that "concept learning" is impossible, because the hypotheses cannot be formulated without already possessing the concept to be learned.

What can be learned through concept learning is *the meaning of words.* A child might learn what the English word "cat" means by forming hypotheses about cats. The child might well form the hypothesis that the word "cat" denotes cats, and then test this hypothesis by observing what people say and do in the presence of cats and other animals. The child may learn in that way that the word "cat" expresses the concept of a cat, but the child does not thereby acquire the concept of a cat. For the child had to have the concept of a cat in order to form hypotheses about cats, or about words meaning cat. Learning the meaning of a word often involves acquiring the concept that it expresses as well as learning that the word expresses it. This is true especially in the case of first-language learners, or students. But it is also common to learn the meaning of a word that expresses a concept that we acquired much earlier, as is the usual case with second-language learners, or code breakers. I believe that the term "concept learning" is a misnomer, based on the widespread but erroneous identification of concepts and meanings (§21.1).

9 Fodor 1998a: 123–4; Fodor et al. 1980: 505; Laurence & Margolis 1999: 10. Contrast Samet & Flanagan 1989: 192–3.

Jackendoff has a different argument for the idea that atomic constituents of thoughts are innate, proceeding from the premise that "[i]n any computational theory, 'learning' can consist only of creating novel combinations of primitives already available" (1989: 98). But there is no reason to insist that such learning is the only means by which concepts can be acquired. Nothing rules out the possibility that the mind-brain system contains a noncomputational process by which perceptual experience produces new primitive conceptual units that expand the scope of a subject's computational processes. That is what I believe abstraction to be.

Fodor concluded in 1975 that concepts must be innate because he thought that concept learning was the only way in which concepts could be acquired from experience, and he realized that a concept could not be acquired as a result of forming hypotheses containing that very concept. His basic premise that concept learning is the only possible method of concept acquisition was groundless, however. Fodor himself called the premise into question, and concluded (1981: 281) that concepts could be acquired more directly, through a "brute-causal" process. But Fodor now considers the direct causation idea dubious on the grounds that it cannot explain why the perception of Φ reliably produces the concept of Φ.

> In short, SIA [a Supplemented version of "Informational Atomism"], like everybody else, has to live with the fact that it's typically acquaintance with doorknobs that leads to getting locked to *doorknobhood*. So, like everybody else, SIA has to explain why it's those experiences, and not others, that eventuate in locking to that property. *But that's enough, all by itself, to make the search for a non-inductivist account of concept acquisition pretty hopeless*. For . . . the hypothesis-testing story has the virtue of solving what I'll call the doorknob/DOORKNOB problem. Why is it so often experiences of doorknobs, and so rarely experience with whipped cream or giraffes, that leads one to lock to doorknobhood? (Fodor 1998a: 127) . . . Why is it the concept X rather than the concept STEREOTYPIC X that one normally gets from experience with stereotypic Xs? (139)[10]

These are very intriguing questions. Their answers would surely provide profound insight into the nature of human beings. But the fact that we do not know *why* or *how* something happens is not an objection to the theory *that* it happens. And the fact that one theory can explain the phenomena while another cannot would not necessarily tip the scales in favor of the former, since theory selection depends on so many factors. Besides, the

10 Cf. Sterelny 1989: 129; 1990: 113, 116; Cowie 1999: 86–9, 99–100.

inductivist theory does *not* explain why experiences of doorknobs cause acquisition of the concept of a doorknob. First, the formation of any hypothesis about doorknobs presupposes prior acquisition of the concept of a doorknob, and thus cannot explain it. Second, the inductivist account raises its own doorknob/DOORKNOB problem. For we can just as well ask, "Why is it so often experiences of doorknobs, and so rarely experience with whipped cream or giraffes, that leads one to formulate hypotheses about doorknobs?" The observation that the experiences constitute evidence for the hypothesis (Fodor 1998a: 127) does not explain the phenomenon, and just prompts the question why it is that evidence for a hypothesis so often causes one to form it. Why is it that we are the sorts of beings who normally respond rationally to evidence?

There are two ways to answer a question like "Why do red objects cause humans to form the concept of red (or to formulate hypotheses about red objects)?" We can specify how the mechanism works, or how humans came to have it. Since the ability to form concepts through abstraction, like the ability to respond rationally to evidence, is an inherited trait, it presumably resulted from evolution, through natural selection. It is plausible that we came to have the ability to respond rationally to evidence because we were better able to survive with it than without it. The same goes for the disposition to acquire the concept of red upon seeing red objects rather than upon hearing blue whales. Surely we would not be as successful a species if perceiving objects in our environment did not give us the concepts that enable us to think about them, and consequently to form beliefs and desires about them. Nor would we be as successful, I should think, if perception produced concepts randomly, with no connection between what is perceived and what is conceived of. The question "How did humans come to have the ability to form the concept of cats in the presence of cats?" seems highly analogous to "How did humans come to have the ability to see cats when cats are present?" No a priori philosophical argument could undermine the answer "Through evolution" to both.[11]

11 In a breathtaking example of armchair science, Fodor (1998a: 129) argues that evolution could not account for such inherited characteristics. Cowie goes further, claiming that "natural selection predicts that concepts should have intentionally arbitrary causes" (1999: 103, also 86–9, 93–4, 111). Fodor's key premise is that "the obvious candidate to select if one wants to ensure that concept acquisition exhibits the d/D relation is inductive learning" (1998a: 129). But inductive learning is not the "obvious" mechanism, as we have seen. And even if it were, all that would follow is that the mechanism of concept formation, like much of nature, is unobvious. See also Fodor 1994: 19–20.

As for how abstraction works, little is known about the process other than the initial and final stages. In the paradigm case, the initial cause is observing Φ, and the final result is acquiring the concept of Φ. In other cases, the initial cause is observing pictures of Φ, or thinking of particular instances of Φ. Even though the details remain to be identified, we do know that the process occurs, and that it is distinct from other processes by which we acquire mental abilities or states. Acquiring the concept of a cat is not properly described as *learning how* to think about or recognize cats, nor as *learning that* anything is true of cats. There is no reason to accept Fodor's premise that concepts are acquired only if they are learned.[12] On the contrary, "learn a concept" – unlike "learn a word" or "learn a fact" – does not seem to be well formed.

It is sometimes said that abstraction involves isolating common features in a series of experiences, and ignoring differences.[13] But a single encounter with a novel animal, for example, often suffices for us to acquire the concept of that species. Moreover, the most salient common feature of cats is that of *being a cat*. It may be the only common feature, if cats are undefinable or share only a set of "family resemblances." But our understanding of abstraction is not increased much by the hypothesis that acquiring the concept of a cat involves isolating the property of being a cat. For that concept too must be acquired by abstraction.

Geach argued that when we observe an object that is colored and red we do not have two sensations, one of red and one of color. He concluded that our concepts of red and color could not have been acquired by isolating the common features of redness and color. There is a reply to this objection: the property of being colored is distinct from the property of being red, and both are observable. We can similarly respond to Humean arguments that the concept of causation could not be acquired by abstraction because there is no sensation or impression of causation; the reply is that we do observe causal sequences and recognize them as causal. But it should be conceded that little is learned from these responses about the mechanics of the process by which we acquire the concepts of red, color, and causation from observation. I hope it is clear that I am using the term "abstraction" as a descriptive term denoting an observed process, not as a theoretical term denoting a hypothetical explanation of the process. It is undeniable that

12 Cf. Samet 1986 and Samet & Flanagan 1989.

13 E.g., Reid 1785: Chapter 5.3; Rand 1969: 15; Benjafield 1992: 63. The view has been criticized in Wittgenstein 1953: 31–2; Geach 1957a: Chapters 6–11; Carruthers 1992: 55; and Gauker 1994: Chapter 1.

children acquire the concepts of red, green, blue, and color by observing red, green, blue, and colored objects. We similarly acquire the concepts of cats, dogs, and animals by observing cats, dogs, and animals. I am calling that process "abstraction" without any implication as to how the process works.

If there is any process at all by which mind and body interact, it has to have a "brute-causal" stage at which a nonmental event causes a mental event directly, with no intermediate mental events.[14] That is, there must be some *basic psychophysical processes*. Plausible examples are the process by which red objects cause sensations of red, or deep cuts pain (sensation); the process by which the intention to raise one's arm causes one to raise one's arm (volition); the process by which anxiety causes perspiration (emotion); and the process by which food deprivation causes a desire for food (appetite). There is no reason to think that we can provide a psychological explanation for any of these processes. On the contrary, it seems that the only way to understand them is to examine their neurophysiological basis and evolutionary origin. We must look at the causal chain that links the responses of retinal receptors to sensations, or intentions to actions. In saying that there is no *psychological* explanation for basic psychophysical processes, I mean only that the causes and intermediate stages of the processes are not themselves psychological events. It does not follow, and is surely false, that basic psychophysical processes will be without any scientific explanation.[15] We have made enormous progress in other areas of psychophysics. While much remains to be discovered, we know a

14 Fodor (1998a: Chapters 6 and 7) asserts in several places that this can hold for formation of the concept of a doorknob only on the assumption that being a doorknob is somehow mind-dependent. Mind-dependence cannot be necessary, however, even on Fodor's own theory, because he acknowledges that the concept of water is acquired directly from experiences of water even though being water is not in his view mind-dependent (1998a: 136). Being a doorknob is mind-dependent, on Fodor's view, because "what it is to be a doorknob is *constituted* by the kind of experience that leads to acquiring the concept DOORKNOB" (1998a: 134). He thinks being red is similarly constituted by appearing red. But this answer is metaphysically unacceptable: doorknobs do not have to cause those experiences any more than red objects have to appear red or water watery. More importantly, Fodor's maneuver just shifts and bifurcates the fundamental question to be answered: why is it that certain experiences and not others cause people to acquire the concept of a doorknob? And why do certain physical objects cause those sorts of experiences and not others? Cf. Cowie (1999: 96–102), who also shows that Fodor's analysis is circular given his view of concept formation as "locking to" properties.

15 Compare and contrast Cowie 1999: vii, 31–2, 60–8, 102, 106–11. Cowie observes that the existence of psychophysical processes would have to be denied if Cartesian or Leibnizian arguments against mind-body interaction were sound, which gives a real bite to their nativism.

great deal about sensation, for example. We know much about light, the eyes, the retina, rods and cones, the optic nerves, the visual cortex, and so on.

It is possible that abstraction is a basic psychophysical process. The causal chain leading from the perception of a cat to formation of the concept of a cat may not have any psychological causes or intermediaries. However, selective attention is surely involved. A child who acquires the concept of a Corvette by observing one must be attending to the Corvette and not to the car's wheels or color or position. Furthermore, there is reason to believe that, like communication, abstraction generally involves some belief formation, or at least some thoughts about the object perceived.[16] It is plausible, for example, that an individual S acquires the concept of a cat when observing some object only if S forms the belief that the object is a distinct kind of animal – that is, an animal different in kind from other animals that S knows. It is similarly plausible that S acquires the concept of a tabby when observing the same object only if S forms the belief that the object is a distinct kind of cat. Then acquisition of "cat" would require prior possession of "animal," and acquisition of "tabby" would presuppose possession of "cat." However, the formation of a concept cannot use the very concept that is to be acquired. Acquisition of "emu" may depend on prior possession of "animal," but cannot possibly depend on prior acquisition of "emu."

The fact that beliefs are generally involved in perceptual abstraction suggests that the process has at least two stages: (1) a perceptual process by which beliefs containing previously acquired concepts are formed about the perceived object; (2) a process of abstractive and/or constructive thought whereby the thoughts activated in stage (1) cause the acquisition of a new concept. The analog of stage (1) for the process of acquiring concepts via communication would be the psycholinguistic process by which thoughts are activated by the speaker's words. Stage (2), it should be noted, is not the process of inference. The input to the process may be a belief, but need not be. And the output is the acquisition of a concept, not the formation of a belief. If perceptual abstraction always has these two stages, then it is not a basic psychophysical process. However, if *all* concept acquisition through abstraction involves belief in this way, then there would have to be some a priori concepts, on pain of an infinite regress. Unless there is an abundant stock of them, it will have to be the

16 Cf. Geach 1957a: 28; Hacking 1975: 64; Sterelny 1989: 129–39; Loar 1991; Fodor 1998a: 124–5; Cowie 1999: 115.

case that acquisition of some concepts does not depend on the prior acquisition of other concepts. So even if abstraction generally involves mediating beliefs, it is an open possibility that in some cases abstraction is a basic psychophysical process. The answer to the question "Why do children acquire the concept of green when observing green objects and not the concept of grue?" may simply be "Human nature." Further understanding may require neurophysiological rather than psychological inquiry. Whether it is basic or not, it seems evident that a complete understanding of abstraction will involve neurophysiological study. The same goes for the processes of abstractive and constructive thought, which appear to be basic *psychological* processes. Concept formation will be more difficult to study than sensation and perception, but there is no reason to rule out similar discoveries about how it works.

This chapter has largely explored issues that do not affect the expression theory of meaning, in order to increase our understanding of one of the central concepts of the theory. One conclusion important for the expression theory is that concept acquisition is distinct from belief acquisition. In particular, concept acquisition is distinct from the process of learning the meaning of words, which I believe is what the traditional term "concept learning" refers to in psychology. There is consequently no circularity or vacuity in the claim that to learn that "red" means "red" is to acquire the belief that the word "red" expresses the concept "red." It is also important to note that postulation of the thought-parts required by the expression theory does not force us to accept incredible forms of the innateness hypothesis. The expression theory is neither an empiricist nor a rationalist theory.

18

The Association of Ideas

We have discussed the fact that ideas or concepts occur to people, are the objects of various propositional attitudes, are parts of other ideas, are expressed by words, are acquired and possessed, represent things, and have contents or objects. Another characteristic property of ideas, one that lends itself to mathematical representation and experimental study, is their *association* with other ideas.[1] We will observe that the association of ideas with sensory images and percepts is an important ingredient in the automatic use and understanding of language. We will also show that neural networks provide a good model and plausible basis for associative networks. But we will examine association mainly because it is a phenomenon distinctive of thought and ideation as opposed to belief and desire. Moreover, it is a causal relationship among ideas distinct from the structural relations obtaining among parts of the same thought. As such, the knowledge that ideas can be associated increases our understanding of the notion of an idea, as does recognition of the fact that ideational structure cannot be reduced to the relation of association.

Historically, philosophers and psychologists who gave ideas a central place in their theories tended to be associationists. Association

1 The subject has a very long history: see Plato, *Phaedo*; Aristotle, *De Memoria*; Descartes, *Treatise of Man*: 90; Hobbes 1651: Chapter 3; Locke 1690: Chapter 2.33; Leibniz 1709: 233; Hume 1739: §1.1.4; Condillac 1746; Hartley 1749; Reid 1785: 455–62; J. Mill 1829: Chapter 3; J. S. Mill 1843: Chapter 6.4; 1869; Bain 1855; James 1890: Chapter 14; Stout 1899: Chapter 11; Wundt 1911: Chapter 3; Titchener 1914; Woodworth 1921: 376–418; Warren 1921; Humphrey 1951: Chapter 1; Price 1953: 317–20; Anderson 1983; Johnson-Laird, Herrmann, & Chaffin 1984; Fodor 1987; Benjafield 1992: 12–15. Experimental studies began in the late nineteenth century with Galton, Wundt, and Ebbinghaus, and would easily fill a library. They are a standard form of research on memory, and are closely related to experiments on conditioned responses. For an introduction to contemporary approaches, see Stillings et al. 1995: 30–2.

psychology was the attempt to analyze and explain all mental phenomena in terms of the association of ideas. This resulted either in ignoring critically important mental phenomena or, more commonly, in applying the term "association" wherever two distinct elements were related in any way, which emptied the term "association" of all useful content. Contemporary connectionists have some of the same tendencies. These limitations of associationism and connectionism should not, however, obscure the importance of the fact that one thought or thought-part may have a tendency to activate another, and to do so through memory according to certain principles. Indeed, contemporary philosophers and cognitive psychologists in their emphasis on "belief-desire" psychology have been equally blind in their neglect of association. The knowledge that particular ideas are associated can explain events that cannot be explained using just the familiar laws relating belief, desire, and action. The purpose of this chapter is therefore both negative – to avoid errors of the past – and positive – to learn more about thought and ideation.

§18.1 ASSOCIATION

Examples of associations abound. A paradigm example of associated ideas is salt and pepper. It is very hard for those of us brought up in America to think of salt without thinking of pepper. The idea of salt is not associated, however, with the idea of singing. Thinking of salt has no tendency to make people think about singing. The idea of snow is associated with the ideas of winter, cold, snowballs, and skiing, but not with the ideas of molluscs, nitrogen, or Saudi Arabia. Thoughts of snow readily elicit thoughts of the former, but not of the latter. Propositional ideas may be associated as well as subpropositional ones. Thus the thought that Edison invented the light bulb may remind someone that Edison invented the phonograph.

Two ideas are associated if *the occurrence of one tends to cause the occurrence of the other in a particular way*. Specifically, the object of one must have a tendency to *remind* a person of the object of the other. Thus salt has a tendency to remind people of pepper, but has no tendency to remind people of singing. Suppose that we are recording S's thoughts, and we find that, over the period of observation, whenever S thought of Jack he thought of Mary. We would be disinclined to attribute this to coincidence. One explanation is that the ideas of Jack and Mary are associated in S's mind. But there may be other explanations. For example, it may be that S's thinking of Jack prompts someone else to utter the word "Mary," which in turn causes S to think of Mary. Alternatively, S may be inferring

facts about Mary from facts about Jack and newly discovered facts about their relationship. In that case, there is a causal connection between the occurrence of the ideas, but not of the right sort for the ideas to be associated. Let us say that idea A *activates* B when an occurrence of A causes an occurrence of B through memory in the characteristic way.

In the usual case, if A has a tendency to activate B, then B also has a tendency to activate A. Thus pepper also has a tendency to remind us of salt. But A may have a tendency to activate B even though B has no tendency to activate A. For example, the idea of a peccary has a strong tendency to activate the idea of a mammal in my mind; but the idea of a mammal has no tendency to activate the idea of a peccary (except in rare situations such as this, in which I happen to have been thinking of peccaries). As long as one of the ideas has a tendency to activate the other, we can say that the two ideas are associated. It is thus true that the idea of a peccary is associated with the idea of a mammal in my mind.

18.1 **Definition:** *Ideas A and B are associated iff A tends to activate B and/or B tends to activate A.*

It follows that association is a symmetric relationship. That is, "A is associated with B" is equivalent to "B is associated with A," and both are equivalent to "A and B are associated." Note well that association is a specific relation *between* ideas, not something we think *about* the ideas.

Definition 18.1 defines association for ideas. English, however, has a convention allowing us to say that any two objects are associated if the ideas of those objects are associated.[2] Let "$\Sigma \rightarrow \Sigma'$" represent the fact that "Σ'" can be used as short for "Σ."

18.3 **Reduction Rule:** *The idea of Φ and the idea of Ψ are associated* $\rightarrow$ *Φ and Ψ are associated.*

These two forms may originally have been used equivalently by mistake owing to the erroneous Cartesian definition of ideas (§19.1). But today the formulas are conventional synonyms. Instead of saying, "The idea of salt is associated with the idea of pepper," we can say more briefly, "Salt is associated with pepper." The latter claim would normally be intended, and understood, as equivalent to the former. Thus interpreted, "Salt is associated with pepper" is a true statement about the association of ideas even though salt and pepper are not themselves ideas, and do not in any sense occur to us.

2 Contrast James 1890: Chapter 14, p. 554.

Associations exist because of causal relations among acts of thought that have certain objects, not in general because of causal relations among the things we are thinking about. When two events are regularly perceived to be causally related – as in the case of lightning and thunder – they will often be associated as a result. But the causal relation between thunder and lightning, which existed before there were any humans, is not the causal relationship in virtue of which the ideas of thunder and lightning are associated in our minds. The ideas themselves have to be causally related. And ideas may be associated even though their objects are not causally related, as the association of salt and pepper illustrates.

Association is a highly relative phenomena, varying with the individual or group, the time, and the context. Ideas associated at one time need not be associated at another. Thus Kennedy and Oswald were not associated until November 22, 1963. Ideas associated in one person or group of persons need not be associated at all or as strongly in another. The ideas of Kennedy and Oswald are associated in the minds of billions of people around the world, but not among the very young or those in isolated cultures. The Fourth of July may be associated with picnics in the minds of Americans, but surely not in those of Tierra del Fuegan natives. There may even be an individual for whom the ideas of salt and singing are associated, even though those ideas are not associated in the rest of the population. Indeed, because I have mentioned salt and singing together for the last few paragraphs, these two ideas are undoubtedly associated to some extent in the reader's mind, as they are now in mine (at least temporarily – the association should weaken rapidly). Because associations differ from individual to individual, psychotherapists examine an individual's associations for diagnostic clues.

Ideas may also be more strongly associated in one context than in another. An important feature of the context in this regard consists of the other ideas that are activated. Thus Britain will be more strongly associated with Wimbledon when discussing tennis tournaments, but more strongly associated with Winston Churchill when discussing World War II. This sort of relativity is to be expected, given the statistical nature of the relation of association. The phenomenon is called *priming* in the cognitive science literature.[3] The negative correlate has been called *inhibition*. Inhibition and priming are the two types of *mental set*. The priming effect, plus

3 See Collins & Loftus 1975: 409; Smith & Medin 1981: 77–8; Fodor 1981: 292–3; Anderson 1983: 96–107; and Stillings et al. 1995: 30, 90–9. Early recognition can be found in Woodworth 1921: 380–5; in the studies of Jost and May summarized in Warren 1921: 219–20; and in the studies of Watt and Ach set out in Humphrey 1951: Chapter 3. Note

our (limited) ability to voluntarily keep a thought occurrent after it has occurred to us, accounts for our ability to *concentrate* and keep a train of thought on a single subject. As long as we keep thinking "What is the best way to design a product?" for example, we will tend to keep thinking thoughts primed by that question, and to avoid those inhibited by it.

The association of two ideas entails not a universal causal law, but only a *tendency* of one to activate the other. The strength of this tendency determines the *strength of association*. Knife and fork are associated ideas, as are knife and spoon. But the former are more strongly associated than the latter, at least for me. For thinking of a knife has a greater tendency to make me think of a fork than of a spoon. In empirical studies, an appropriate measure of correlation can be used as a measure of association strength. A simple measure would be the average of the *activation probabilities*. The idea of a knife will have a certain probability of activating the idea of a fork, and the idea of a fork will have a certain probability of activating the idea of a knife. The degree to which knife and fork are associated in S's mind would then be determined by the average of these probabilities. Another measure of association strength is *reaction time* or *latency*, the time elapsed between the occurrence of one idea and the occurrence of another. After I have thought of a knife, I will generally think of a fork before I think of a spoon.

The *laws of association* specify variables that influence association strength. Two of the most familiar are *frequency* and *recency*. Other things being equal, the more frequently two ideas occur together, the stronger their association tends to be. In the extreme, whole trains of thought may even become habitual. The frequency law explains, for example, why salt and pepper are more strongly associated than salt and singing. Similarly, the more recently two ideas have occurred together, the stronger their association. Suppose that our friend Bill was married to Jill for ten years, but has been married to Jane for the last ten years. Even if we have thought of Bill and Jill together as many times as we have thought of Bill and Jane together, we should have a greater tendency now to be reminded of Jane when thinking of Bill than to be reminded of Jill. For the ideas of Bill and Jane have occurred together more recently. It should be clear that such laws of association are laws of thought rather than of belief (§12.2).

that a proposition, or even an entire "script" (Stillings et al. 1995: 33–5) or "regular train of thought" (Reid 1785: 446–54; Warren 1921: 66) may also be primed. James (1890: Chapter 14, pp. 567–9, 583–94) was most insightful, anticipating most of what I shall say.

The *internal* form of the laws of association just discussed relates the strength of association to the co-occurrence of the two ideas. The *external* form relates the association of two ideas to the co-occurrence of their objects. The external form holds as well as the internal in the special case in which the objects are regularly recognized. In that case, the co-occurrence of the external objects causes the co-occurrence of the ideas of those objects. The laws of association then say that the strength of the association between two ideas is increased whenever the objects of the ideas co-occur. The external form of the laws of association thus implies that the associations among our ideas are partially controlled by the correlations of external objects. In one respect, our mind becomes an analogue model of the external world. The ideas of thunder and lightning occur together in our minds in part because thunder and lightning occur together in nature. The ideas of salt and singing do not tend to occur together in our minds because salt and singing do not often go together in reality. The standard *paired associates* or *verbal learning* experiment exploits the external form of the laws of association. Thousands of experiments have been conducted in which the ability of a subject to recall one stimulus, S_1, when presented with a second, S_2, is treated as the dependent variable, and the recency or frequency of pairings of S_1 and S_2 is taken as an independent variable, or is controlled to study some other factor.[4]

The fact that two ideas are associated represents a general causal tendency that we can use to explain particular occurrences.

> Sir William Hamilton relates, for instance, that after thinking of Ben Lomond he found himself thinking of the Prussian system of education, and discovered that the links of association were a German gentleman whom he met on Ben Lomond, Germany, etc. The interesting part of Ben Lomond, as he had experienced it, the part operative in determining the train of his ideas was the complex image of a particular man. (James 1890: Chapter 14, p. 578)

In this classic case, what is to be explained is why Hamilton thought of the Prussian system of education on a particular occasion, right after he thought of Ben Lomond. This is a curious happening, whose explanation is far from obvious. The explanation that Hamilton provided cites two key associations: the idea of Lake Lomond is associated in his mind with the idea (or image) of a certain German gentleman, and the idea of that German is associated in his mind with the idea of the Prussian

4 See, e.g., Krech, Crutchfield, & Livson 1969: Chapter 21; Marx 1969: Part V. The locus classicus is Ebbinghaus 1885.

system of education. The first association is explained in turn by the co-occurrence of those two ideas on a recent trip, and the second association is explained by the frequent co-occurrence of the ideas of Germany and Prussia resulting from the extensive historical relation between Germany and Prussia.[5]

Standard "belief-desire psychology" is as incapable of explaining Hamilton's train of thought as "associationist psychology" was incapable of explaining why people act and reason as they do. What is to be explained is not the formation of a belief or desire, nor the performance of an intentional action. Hamilton's thinking of the Prussian system of education was not the product of theoretical or practical reasoning.

Many rules of "natural deduction" used in standard formulations of logic have analogues that are widely applicable psychological tendency laws. Thus Modus Ponens has a psychological as well as a logical form.

(1) *Logical Rule*: From "p" and "if p then q," "q" can be validly inferred.
Psychological Rule: People who believe "p" and "if p then q" tend to infer "q."

The psychological rule can be used to explain why Kathy believes that John was playing in the woods: she believes that John has poison ivy; she believes that he must have been playing in the woods if he has poison ivy; and she has inferred in accordance with Modus Ponens. But Kathy would not have made this inference if its premises had not occurred to her. To provide a complete explanation of her belief that John was in the woods, we have to explain why it occurred to Kathy that John has poison ivy and that he must have been in the woods if he has poison ivy. It is not enough to attribute to Kathy further beliefs, such as that John has poison ivy if he has a particular type of rash. For people may have beliefs without their occurring on a particular occasion. What we need to know is why a certain idea occurred to Kathy at a particular time. The fact that Kathy associates poison ivy with that type of rash, and with the woods, may well provide part of the missing explanation. Perception may also play a role in the explanation, of course. Kathy may have perceived that John had poison ivy. Even if she did, that fact too would imply that Kathy associates poison ivy with a certain sort of rash.

We can use our knowledge of associations to predict and control behavior. As a trivial but obvious example, if we realize that Bill has forgotten to buy pepper, we can help by saying "Salt!" More seriously, if we want to

5 Associational explanations for otherwise surprising sequences of thought were also noted by Hobbes (1651: Chapter 3).

keep someone in a good mood, it is often necessary not only to avoid talking about certain sensitive subjects, but also to avoid mentioning things associated with those subjects. We often help students to remember the answer to a question by dropping hints. Marketers make it their business to expand and exploit their knowledge of standard associations. Writers select words carefully for their connotations, which are things associated with, but not meant by, the words. Education is important not only for instilling a body of facts, but also for creating a set of associations whereby relevant facts are recalled to mind when they are useful. A trader may know the mathematical rule for calculating the volume of a cylinder. But unless he associates that rule with questions about the quantity of product in a barrel, his knowledge will be useless in helping him to determine how much he should pay or charge for it.

Association psychology was transformed by the behaviorist revolution in the first half of the twentieth century. The behaviorists Watson, Skinner, and Hull, inspired by Pavlov and Thorndike, focused on *conditioned responses*.[6] In *classical* conditioning, the experimenter selects an "unconditioned stimulus" (UCS) that automatically produces a reflexive "unconditioned response" (UCR). The experimenter then pairs the UCS with a "conditioned stimulus" (CS), which does not initially produce UCR. The typical finding is that after the subject repeatedly perceives the UCS and the CS together, the CS acquires the ability to evoke the same response, now called the "conditioned response" (CR). Thus Pavlov's dogs initially salivated (UCR) when presented with food (UCS). After repeatedly hearing the bell (CS) when presented the food, the dogs began salivating (CR) when the bell was rung, before the food was presented. The experimental result can be explained in part by hypothesizing that the subject forms an association between the ideas of the UCS and the CS. The mere belief or knowledge that the CS precedes the UCS is not sufficient to explain the conditioned response, for that cognitive state may be present when the subject is not thinking about either stimulus. Furthermore, the motivational principles of belief-desire psychology do not apply, because the response is involuntary. In order to explain why the conditioned response occurs when the CS is presented, we need to hypothesize that the subject is thinking of the UCS, which in turn is explained by its association with the CS. Graham's (1988: 154) argument that no *representation* of the CS is involved, because no *inference* or *rule-following* is involved, epitomizes

6 Thorndike 1911; Watson 1924; 1930; Pavlov 1927; Skinner 1938; 1957; Hull 1943; Kimble 1961.

how much cognitive psychology has been shackled by its narrow focus on belief and desire and its neglect of thought.

In *instrumental* or *operant* conditioning, the experimenter selects a voluntary response, R, a discriminative stimulus, S_d, and a reinforcement, S_r. In the classic experiments of Skinner (1938), R is pressing a bar, S_d is a light, and S_r is a food pellet. The experimenter presents the reinforcement when the subject performs the response in the presence of the stimulus, and withholds the reinforcement otherwise. The typical finding is that after a learning period, the subject reliably performs R at high rates when S_d is present, and seldom if ever without S_d. The principles of belief-desire psychology are clearly involved in explaining this result. The subject must learn that the response produces the reward only if the stimulus is present. This belief acts together with the subject's desire for the reinforcement to produce the response. But equally important in explaining the result is the fact that the subject associates the idea of S_d with the ideas of R and S_r. For it is this association that explains why the subject thinks of R and S_r when S_d is present, and why the subject recalls that the performance of R produces S_r when S_d is present.

The associative connection of ideas with impressions and images is essential in the learning and use of language. Whereas a word like "cat" gets its meaning because it is conventionally used to express the idea of a cat (Chapter 8), learning to use and understand the word effortlessly requires acquiring the ability to recognize the word "cat" automatically, and developing a strong association between the idea of the word and the idea expressed by the word. The laws of association imply, as a special case, that the more frequently or recently a word is used or interpreted in a given sense, the more likely it is to call up the corresponding idea. The phenomenon of priming helps us to cope with ambiguity. The word "tank" automatically calls up the idea of an armored self-propelled combat vehicle when I am reading about wars, whereas the idea of a fish tank is activated when I am reading about tropical fish. Complete mastery of a language requires knowledge of the connotations of words as well as their meanings. A word "w" meaning μ connotes ν only if the idea of the word is associated with the idea ν as well as the idea μ. Normally, the ideas themselves will be associated with each other. For obvious reasons, the name "Rolls Royce" connotes wealth and luxury, while "K-Mart" connotes low-class and cheap.

I am not proposing that the laws of association are fundamental principles of psychology, nor that all mental phenomena can be accounted for as the modification of sensations by association. Most of the theoretically

ambitious claims distinctive of associationism are false.[7] The classical laws of association are true, but they have limited predictive power. Given all the ideas occurring to us at any given time, for example, it is hard to predict which of them will become associated. Moreover, the fact that two ideas are associated in a given individual's mind, as well as the general mechanism of association, may cry out for explanation. If everyday examples do not convince you, think of psychoanalytic cases. Furthermore, the concept of association is not very fundamental conceptually. Far from being an alternative to the faculty of memory, or something in terms of which memory can be defined, I believe that association must be defined in terms of memory. And for reasons that should be abundantly clear by now, the meaning of a word cannot be defined as "the associated idea" (§8.1). None of these caveats, however, lessens the importance of association as a psychological phenomenon. A narrow cognitivism focusing exclusively on belief and desire suffers from explanatory and descriptive incompleteness just as surely, if not as massively, as associationism.

§18.2 ASSOCIATIONIST NETWORKS

Associated ideas form *networks*, which can be represented by network maps. In associative network maps, nodes represent ideas, arrows represent activation, and the numerical measures represent activation probabilities.[8] A simple example is provided in Figure 18.1, which represents a network of three ideas: *Animal*, *Cat*, and *Dog*. Each node is connected with two arrows, since the probability of one idea activating a second need not equal the probability of the second activating the first. Thus Figure 18.1 represents the probability of *Animal* activating *Cat* as .2, and the probability of *Cat* activating *Animal* as 1.0. This would reflect the fact that it is practically impossible for the subject (or group of subjects) S to think of a cat without thinking of an animal, but it is common for S to think of animals without thinking about cats. Figure 18.1 represents

7 For examples, see Condillac 1746; Hartley 1749; J. Mill 1829; James 1890: Chapter 14, pp. 598–604; Titchener 1914: esp. Chapters 7 and 9; Warren 1921: Chapters 3, 6, 9, 10; and Fodor 1987: 139. Contrast James 1890: Chapter 1, pp. 1–3, Chapter 14, p. 565; Wundt: 1911: Chapter 3; and Humphrey 1951: Chapter 1.

8 See, e.g., Smolensky 1988: 28; Fodor & Pylyshyn 1988: 102ff.; C. MacDonald 1995: 9. In some theories, "activation" denotes a hypothetical quantity (much like current), the amount of which determines whether an idea occurs in conscious experience. I would call it "activation *energy*" or "activation *potential*." The numerical weights determine how much activation potential spreads along a given pathway. This model provides a theoretical explanation of the activation probabilities I have represented.

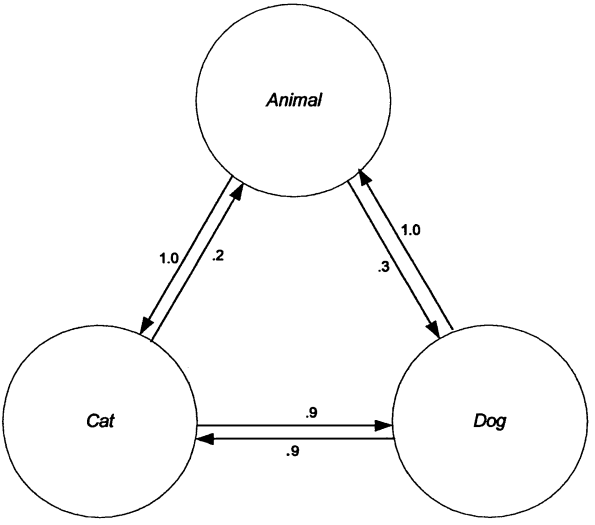


Figure 18.1. Associative network

Cat and *Dog* as being more strongly associated for S than *Cat* and *Animal*, despite the fact that neither *Cat* nor *Dog* invariably activates the other.

A complete associative network map for any given person would be staggeringly complex. There would be literally millions of nodes, each of which is connected to hundreds or even thousands of other nodes. For a vivid model of the stream of consciousness, imagine an enormous electrical network of blinking lights. When a light is on, it represents the occurrence of an idea. Add that each light causes connected lights to come on with the specified probabilities, and the model is complete. It is complete for "free thought" – that is, thought that is free of nonideational influences, such as sense perception or volition. A more interesting and realistic model is provided by a live neural network, with neural structures representing ideas, and the activation of a structure representing the occurrence of an idea. We will explore this idea more fully in §18.4.

We observed in §18.1 that some associations may be primed and others inhibited by the occurrence of further ideas. Figure 18.2 represents two situations, one in which the idea of World War II is activated (shading), and one in which it is not. The network map in Figure 18.2 represents *Britain* as more likely to activate *Winston Churchill* when S is thinking of World War II than otherwise.

The fact that two ideas are associated does not tell us what S's attitude is toward any proposition containing those two ideas – whether S believes

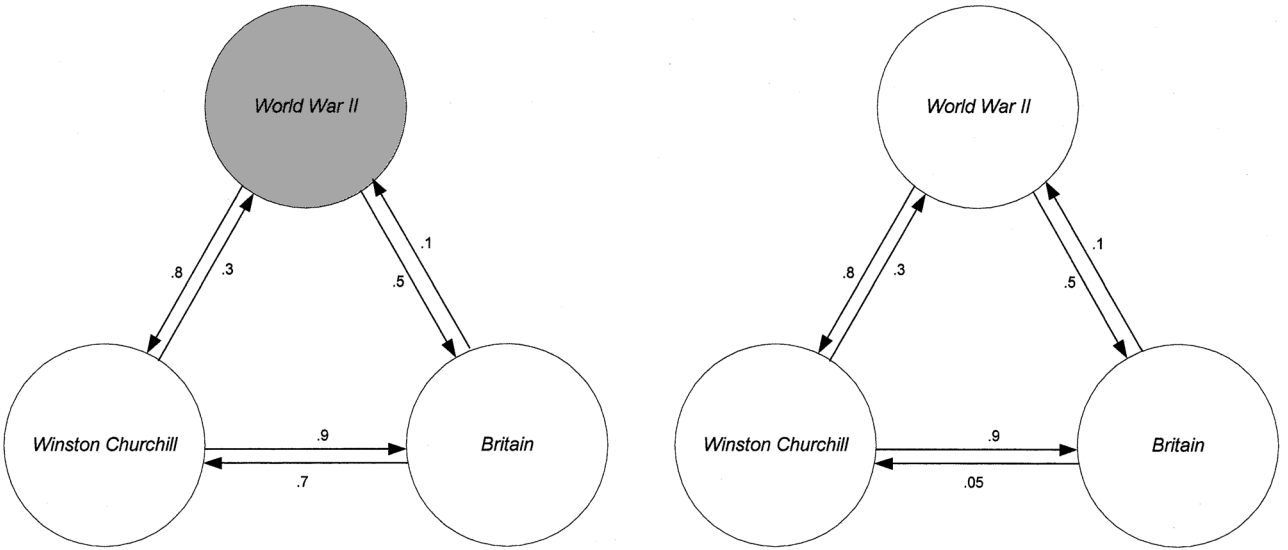


Figure 18.2. Priming effects

or desires it, for example.[9] The numerical measures in association maps like Figure 18.1 do not represent degrees of belief. S will probably be completely certain ($b = 1$) that both cats and dogs are animals, while having no belief at all that dogs are cats ($b = 0$). Most people are as sure that cats are not dogs as they are that cats are not roses, despite the fact that cats are associated with dogs but not with roses. The fact that concept "A" is associated with concept "B" does not predict that S believes that A is B (think of "salt" and "pepper"), nor that S believes that A is not B (think of "Bush" and "the president"), nor that S believes that A is similar to B ("diameter," "π"), and so on.

§18.3 ASSOCIATION VERSUS CONSTITUENCY

Of special importance to us is the fact that the phenomenon of association cannot explain *ideational structure*, and is not a *combinatorial* or *uniting* operation.[10] The fact that S associates the light bulb with Edison does not entail that S is conceiving, or has a tendency to conceive, any complex idea containing the simpler ideas "Edison" and "light bulb." Still less does that fact tell us what sort of complex idea is occurring to S if those two ideas are occurring together in some complex. It does not tell us, for example, whether S is thinking "Edison invented the light bulb," "Edison is wrongly credited with having invented the light bulb," "The light bulb invented by Edison had a tungsten filament," or something else. Conversely, the fact that S is thinking of purple spaghetti entails that the ideas "purple" and "spaghetti" are occurring together in S's mind, but does not entail that those ideas are associated in S's mind. The more critical associationists distinguished "simultaneous association" (co-occurrence as part of a single complex idea) from "sequential association" (a tendency for an occurrence of one to cause the occurrence of the other). But the two relationships have little in common. By "association" we mean only "sequential association," as Definition 18.1 makes clear.

9 Contrast Quillian 1968: 227; Rey 1991: 229; Ramsey, Stich, & Garon 1991: 209.

10 Cf. Stout 1899: 431; Humphrey 1951: 13, 72–3; Fodor & Pylyshyn 1988: 102–3. Contrast Hartley 1749: Proposition 12; J. Mill 1829: 114–15; J. S. Mill 1869: 164; Bain 1855: 2.4; Titchener 1914: Chapter 9; Warren 1921: 9–13, 38–9, 54–5, 65–6, 87, 162–6, 260–5, 276–7, 287–8, 306; Quillian 1968: 223, 228–9; Katz 1966: 240–1. Contrast even Fodor (1983: 27), who said once that a central tenet of classical associationism is that "the property of being associable is preserved under association, . . . thereby generating a distinction between elementary psychological structures and complex ones." He adds that what modern logic and associationism have in common is "the idea of sets of elements with combinatorial operations specified over them" (1983: 30).

Some network models perpetuate the associationist error of overextending the concept of association. Consider the treatment of "propositional networks as associative networks," described in Stillings and associates (1995: 27–32).[11]

> We will informally introduce a simple version of *propositional network* notation, which involves the theoretical commitments that are most important in cognitive psychology . . . . In this notation each proposition is a unique structural unit built around a *propositional node*, which is symbolized by a circle with a number inside it. The value of the number has no meaning other than to distinguish the proposition from other propositions. The propositional node is connected to its *constituent* nodes by *links*, or *pointers*, which are symbolized by arrows. Each link has a label that specifies the role that its target plays in the proposition. Every propositional node must have a link to a relation node, which is symbolized by a word in capital letters. The propositional node must also have the right number of links to nodes for the arguments of the relation. (Stillings et al. 1995: 27)

Their example of a propositional network appears in Figure 18.3, in which four propositions are represented: (1) *Mary likes John*; (2) *A is an apple*; (3) *John is a teacher*; (4) *John gave A to Mary*. As a means of representing propositions, this symbolism has limitations. It is hard to see, for example, how propositions with more than one quantifier could be represented; consider the ambiguity of "Someone loves everyone." But the problem I wish to point out arises from treating the relationship between a proposition and its constituents as an associative connection.

> A number of further assumptions must be introduced to make propositional networks a full-fledged theory, or *model*, for long-term declarative memory. . . . The basic assumption of the theory is that at each moment in time each node in a network is at some level of *activation* and that activation spreads among nodes along the links. If the level of activation reaches a high enough value in some portion of the network, that portion of the network is accessible to conscious attention. The links are therefore *associative* connections that determine the tendency of one item of information to lead to another in thought. (Stillings et al. 1995: 30)

This theory embodies several fundamental confusions. First, as argued in §14.2, a proposition is a complex *composed of* its component concepts, not a separate idea that is *associated with* or *activates* those concepts. If S is thinking *Mary kissed John* and also *John kissed Jane*, it follows that the ideas

11 Stillings and colleagues are here describing the approach developed by Quillian (1968), Collins & Loftus (1975), and Anderson (1983), and critiqued by Johnson-Laird, Herrmann, & Chaffin (1984). Contrast Fodor & Pylyshyn 1988: 102–11.

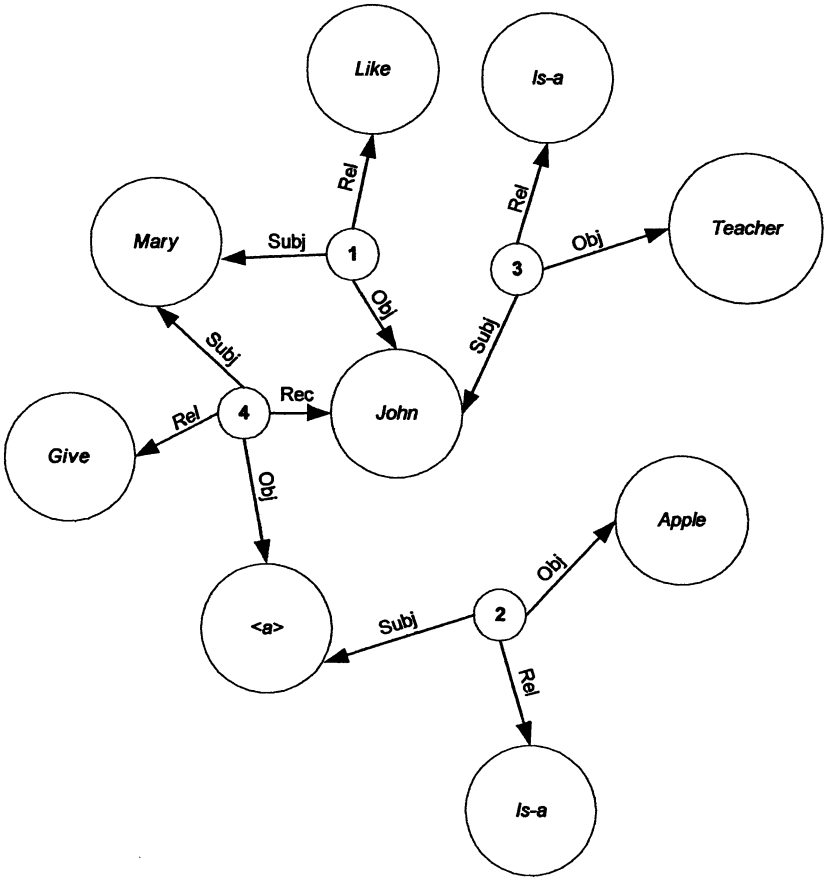


Figure 18.3. Propositional network

Jane, *kissed*, and *Mary* are all activated. But even if these three ideas are strongly associated, it does not follow that S is thinking *Jane kissed Mary*. The ideas have to be occurring in the right structural relationships for that thought to occur.

Second, propositional network theory would allow propositional node ① to be activated sufficiently for conscious attention without the constituent nodes *John* and *Mary* being activated sufficiently for conscious attention. This would imply, absurdly, that the proposition *Mary likes John* could be occurring in consciousness without the concepts *Mary* and *John* occurring in consciousness. Third, the theory would appear to imply erroneously that any idea that is associated with or activates *Mary likes John*

must also be associated with and activate *John likes Mary*. Since anyone who can think the first thought can think and maybe even has thought the second, both must be represented in the propositional network. So anything linked to the one will be linked to the other.

Finally, as the talk of "declarative memory" and "information" indicates, propositional networks are intended to model remembering *that* p, which implies *believing* that p. However, association concerns thought rather than belief.

§18.4 CONNECTIONIST MODELS

A *connectionist model* is a hypothetical network, modeled on the structure of the brain, that is designed to explain or predict the associative relations among a set of cognitive elements and the dynamics of their association.[12] A connectionist model postulates a network of units or nodes, each of which has excitatory and inhibitory connections to other units. The activation of one unit has a probability of activating other units that is determined by the connections among the nodes. The probabilities can change over time, as connections are strengthened or weakened. The nodes are modeled on neurons or neural structures, with their excitatory and inhibitory synaptic connections to other neurons. The activation of one neural structure is known to cause the activation of physically connected structures with specifiable probabilities that are directly related to the number of excitatory connections and inversely related to the number of inhibitory connections.

We noted in §18.2 that a plausible model of the association of ideas can be provided by letting the activation of a neural structure represent the occurrence of an idea.[13] We could hypothesize that an idea is the pattern of activation displayed by a particular neural structure, and that to have the concept is to have the neural structure in operational condition. The association of two ideas might then be explained in terms of the synaptic connections among the neural structures involved. The networks in the brain seem to be at least as complex as those among associated ideas. The synaptic connections are also known to change with development and learning, and to decrease with disuse. Neural structures tend to develop

12 For classic sources, see Rumelhart & McClelland 1986; Horgan & Tienson 1991; and Macdonald & Macdonald 1995.

13 We could not identify ideas with the neural structures themselves, since they are of the wrong ontological type. Ideas are event-types, while neural structures are three-dimensional material objects.

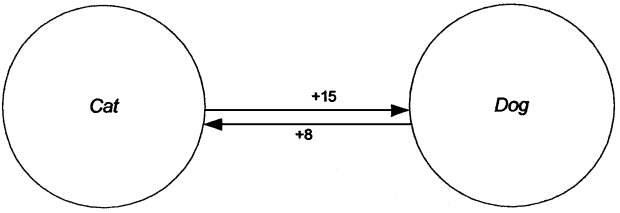


Figure 18.4. Connectionist network

synaptic connections in proportion to how often they fire together. These facts provide a plausible basis for the familiar law relating recency and frequency of co-occurrence to association strength.

In neural networks, the synaptic connections responsible for causal interactions can be either excitatory or inhibitory. Let n(ι) designate the neural structure whose activation is idea ι. It is natural to suspect that the degree to which idea *A* tends to activate idea *B* is directly related to the number of excitatory connections that n(*A*) has with n(*B*), and inversely related to the number of inhibitory connections. In connectionist models, it is customary to abstract from this level of detail and to represent the net excitatory or inhibitory effect with a single positive or negative number, called a strength or weight. In Smolensky's (1988) models, the influence of unit *i* on unit *j* is the activation level of *i* times the strength of the connection from *i* to *j*. The influence of *i* on *j* is excitatory if the connection strength is positive, inhibitory if the weight is negative. The activation level of a node is a function of the influence of all of the nodes connected to it. Hence the probability that *i* will activate *j* is determined not just by the strength of the connection between *i* and *j*, but also by the strengths of all of the other connections to *j*. Thus whereas the associative network in Figure 18.1 represents *Cat* and *Dog* as equally likely to activate each other, the connectionist model in Figure 18.4 represents the connections between them has having different strengths. The two figures are fully compatible, because the probability of *Cat* activating *Dog* is determined not just by the influence of *Cat* on *Dog*, but also by the influence of *Cat* on all of the other ideas in the network that influence *Dog*. In general, it is important to distinguish maps such as Figure 18.1 that seek merely to describe the observed facts of association, from theoretical maps such as Figure 18.4 that attempt to represent the underlying causal processes.

The failure of association psychology implies that not all relations among ideas can be explained by excitatory or inhibitory connections among activation patterns representing those ideas. Some other factors

would have to be introduced to account for what is believed and desired as opposed to what is merely conceived. Furthermore, mereological connections among ideas would have to be identified with the mereological connections among the activation patterns that they are identified with. That is, if idea *A* is part of idea *B*, then the activation pattern identified with *A* would have to be part of the activation pattern identified with *B*. It would follow that the probability of *A* occurring given *B* is 1, but this would not be the case where the occurrence of *B activated* the occurrence of *A*, nor where *A* occurred because of excitatory stimulation from *B*. Failure to recognize the difference I am pointing to will needlessly restrict connectionist models. Since neurons are parts of larger neural structures – the largest being the various lobes and the brain itself – and since larger neural structures can also have synaptic connections with each other, there is no reason why connectionist models have to focus exclusively on networks of *simple* elements, nor why such models have to explain all relations among ideas in terms of synaptic connections rather than in terms of parthood or other features.

There is also no reason why connectionist models of propositions and propositional attitudes must restrict themselves to nodes representing only the propositions and their components. We observed earlier that to be thinking the thought *Mary likes John*, the component concepts *Mary*, *likes*, and *John* must be occurring to the subject in certain relationships, and that the necessary relationship is not that of association or mere coactivation. We could postulate that the relationship involves being coactivated by a distinct type of neural structure, which we might call a *thought connector*. That is, we could hypothesize that the thought that Mary likes John occurs to a subject only if a thought connector coactivates the ideas *Mary*, *likes*, and *John*. (To account for the difference between thinking that John likes Mary and thinking that Mary likes John, we would also have to postulate that thought connectors can activate these concepts in two different ways.) We could then reinterpret Figure 18.3 by taking nodes ① through ④ to be four different thought connectors rather than thoughts. The thought that Mary likes John would be a whole consisting entirely of the concepts represented by *Mary*, *John*, and *likes*. The point would be that the three occurrences of these concepts comprise the thought that Mary likes John only if they were caused in the right way, namely, by a separate neural structure. These relationships are clarified in Figure 18.5, in which the region labeled *Ljm*, rather than node ①, represents the thought that John likes Mary. On this reinterpretation of Figure 18.3, the arrows do not represent *associative* connections, and the diagram is not a model of

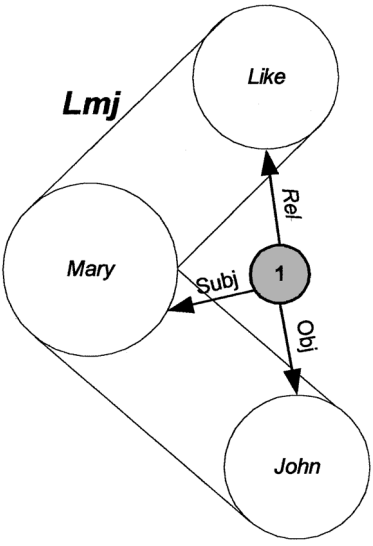


Figure 18.5. Thought connector

memory of any sort. It is only a model of occurrent thought. People can think thoughts involving concepts that are not associated in their minds. These thoughts may not be stored in memory in any way, and may never reoccur.

We could turn the connectionist model of thought presented in Figure 18.5 into a connectionist model of belief by connecting the ideational network to a system of nonideational nodes that we might call the *attitude network*. Let the attitude network contain one or more *belief nodes.* For this to be a plausible model of belief, we have to suppose that the process of learning that Mary likes John produces changes in the excitatory and inhibitory connections among the nodes within the attitude network, with the result that after the learning process, activation of the thought "John likes Mary" activates a belief node.[14] The belief that John likes Mary could then be identified with the disposition of the attitude network to produce an activated belief node when given an occurrence of the thought that John likes Mary. Occurrently believing that John likes Mary could be identified with a structure of activation in which the thought actually

14 This model was inspired by the connectionist models of memory represented by networks A and B in Ramsey, Stich, & Garon 1991. They believe, though, that their models are incompatible with commonsense psychology, and conclude that cognitive science can do without representations or propositional attitudes.

activates a belief node in this way. Degree of belief might be identified with some function of the level of activation of the belief node. And disbelief could be inhibition of the connected belief node.

If the attitude network represented all propositional attitudes, it would have to contain nodes representing desire and all other fundamental propositional attitudes. Derived propositional attitudes could be represented by patterns of activation in the fundamental nodes. Being optimistic that John likes Mary might be identified with the disposition of the attitude network to activate both belief and desire nodes when the thought that John likes Mary is the input.[15] It is important to note for present purposes that the belief and desire nodes are not ideational nodes, and thus differ fundamentally from the *John* and *Mary* nodes. The activation of a belief node does not represent the subject's thinking about belief or anything else. It is a nonrepresentational node.

In conclusion, since connectionist models postulate a network of nodes with excitatory and inhibitory connections, they provide a natural model for systems of associated ideas. But connectionist models can also provide models of nonassociative cognitive phenomena, such as thought and belief. Since connectionist networks are themselves modeled on the structure and function of the brain, the ability to provide connectionist models shows how psychological phenomena could be physically realized.

15 See Davis 1981a; 1981b; 1984a; 1987; and 1988b for the sort of analytic work this idea presupposes.

19

Objects, Images, and Conceptions

I argued in Parts I and II that the venerable formula "meaning consists in the expression of ideas" is true for the vast majority of expressions when the term "idea" denotes thoughts and their parts. I have devoted Part III to clarifying the notion of thought and ideation invoked in the expression theory, defending the assumption that thoughts have parts, and showing that occurrent thought plays an essential and distinctive role in the explanation of human behavior. I believe that the definition of "idea" and "concept" as denoting thoughts and their cognitive parts (Definition 15.1) is an accurate analysis of one of the standard meanings of these terms rather than an unconventional stipulation. Nonetheless, ever since the "new way of ideas" emerged with Hobbes, Descartes, and Locke, the terms "idea" and "concept" have been officially applied by philosophers, psychologists, and linguists to almost everything but thought-parts, including objects of thought, contents, images, words, inner speech, and conceptions or belief systems. Some of this usage seems to have been motivated by reductive theorizing, some by simple blindness. Nearly everyone who has used the terms "concept" and "idea" has done so inconsistently and unrigorously, defining them differently on different occasions or applying them when their official definition does not apply.

A man is thinking of the sky, on my analysis, provided a certain thought-part is occurring to him, namely, the idea of the sky. In general: *S is thinking of Φ iff the idea of Φ is occurring to S* (Definition 12.2). With or without the background conception of thought-parts, this connection between thinking of Φ and the idea (or concept) of Φ has been accepted explicitly or implicitly by nearly everyone who has spoken of ideas, and is extremely well entrenched. Its acceptance, I believe, is what has led those who define "idea" differently into inconsistency. The equivalence

we are using as Definition 12.2 would have to be abandoned if ideas were defined as anything but thought-parts.

Since our primary objective is the formulation of an acceptable theory of meaning, it is imperative that we make it crystal clear how thought parts differ from other things that may go by the same name, and with which thought-parts might be confused. We will gain a better understanding of thought-parts in the process, by detailing how they resemble and differ from closely related entities. Seeing what something is *not* is an important part of knowing what it is. We also gain a deeper understanding of occurrent thought as a propositional attitude distinct from belief, which has an important but neglected role to play in cognitive science.

§19.1 IDEAS VERSUS OBJECTS OF THOUGHT

Definition 12.2 does *not* say that we think only about ideas, nor that we cannot think about Mary without thinking about the idea of Mary.[1] Thinking about the idea of Mary requires the occurrence of a second-order idea, the idea of the idea of Mary. Thinking about Mary requires the occurrence of a first-order idea, and does not require thinking about that idea. Ideas are not things that we think about in order to think about the world. Ideas are things whose occurrence *constitutes* thinking about the world. What we think *about* are the *objects* of the ideas in our thoughts.

Descartes, however, and following him Arnauld, Locke, and Leibniz, often defined ideas as *objects* rather than *parts* of thoughts, meaning *intentional* objects (§12.4). That is, they defined an idea as *what a thought is of* or *about*.

(1) An idea is an (intentional) object of thought (i.e., what a thought is of or about).[2]

1 Russell (1910–11: 222) correctly opposed making ideas a "veil" between us and the real world, but mistakenly took this to be a consequence of the thesis that ideas are parts of judgments. See also Geach 1957a: 59–60; Hill 1991: 67; Kapitan 1994: 283.

2 Descartes 1641b: 9–10, 52; Arnauld 1641: 87, 106; 1683: Chapters 5–7; Locke 1690: Introduction, §8; §2.1.1; Leibniz 1709: §2.1. See also Berkeley 1710: §§1.7, 1.38, 1.89, 1.140, 1.142; Joseph 1916: 21–2; Humphrey 1951: 274, 315; Urmson 1967: 119; Burge 1979b: 537; and Bradshaw 1991: 422–3. Contrast Gassendi 1641: 157; Reid 1785: §2.4.2; James 1890: 461; and Husserl 1900: 355. Elsewhere, Burge (1979a: 425) defines concepts as thought-parts, without indicating that there is any significant difference. Ockham held that universal concepts were thought objects in his *Ordinatio* (1320: 44), but rejected the view in his later works (see Boehner 1964: xxix). Leibniz elsewhere states that "the idea of a circle is not exactly like the circle" (1676: 283). Descartes himself defines "idea" quite differently on different occasions (e.g., 1641b: 52; 1641a: 159–60), and adds a qualification to his definition

I will refer to (1) as the *Cartesian definition*, although Descartes sometimes defined the term differently. Few definitions have been so observed in the breach, and for good reason. Since the Cartesian definition appears to lead straightaway to metaphysical idealism, its acceptance would require abandoning the most fundamental and well-supported beliefs about both physical objects and ideas. Right now I am thinking of the Sun. The Sun is therefore the intentional object of my thought. By the letter of the Cartesian definition, the Sun would therefore be an idea. But the Sun is a massive physical object about 93 million miles away from here. Since the Sun is composed of an incredible number of hydrogen atoms, the Cartesian definition implies that the idea of the Sun is composed of hydrogen atoms.[3] As Reid argued, it is absurd to say that an idea has such properties. It also follows from the Cartesian definition that containing hydrogen is the same thing as containing the idea of hydrogen. It is just as absurd to say that the Sun contains ideas as it is to say that ideas contain hydrogen. Given (1), we would have to decide whether the Sun was nonexistent before human beings evolved and acquired the idea of the Sun, or whether the idea of the Sun existed billions of years before cognitive processes did, because the Sun is over four billion years old. When the object of thought is itself a mental event, we get a different sort of absurdity. Since I cannot think of pain unless the idea of pain occurs to me, the Cartesian definition implies that I cannot think of pain without experiencing pain, which is demonstrably false. One might try holding that there are two Suns, the idea in my mind and the star in the sky. But when I am thinking of the Sun, as I often do, I am thinking of the object

that we will discuss later. In the *Treatise on Man*, ideas are even taken to be "currents in the spirits" in the brain. See Urmson 1967 for a history of the term "idea."

It should be noted that following Descartes (e.g. 1644: pr. 1.9), Locke generally took "thought" to include not only thinking in our narrow sense, but all propositional attitudes, plus perception, willing, feeling – in short, "all that of which we are conscious as operating in us."

3 This sort of inference was explicitly accepted by Leibniz: "It is obvious that green, for instance, comes from a mixture of blue and yellow; which makes it credible that the idea of green is composed of the ideas of those two colours, although the idea of green appears to us as simple as that of blue, or as that of warmth" (1709: §2.2.1). The same confusion is found in current authors: "Suppose that a subject were asked to state everything he knows *about the concept 'machine.'* . . . This information will start off with the more 'compelling' facts *about machines*, such as that they are usually man-made, involve moving parts, and so on" (Quillian 1968: 227; my emphasis). Cf. Leibniz 1676: 283; Berkeley 1710: §1.38; J. Mill 1829: 264; and Marconi 1990. Contrast Reid 1764; 1785: 488–9; James 1890: Chapter 9, pp. 236, 277; Husserl 1900: §4.2; Brody 1969: xx–xxi; Collins & Loftus 1975: 408–10; and Kapitan 1994: 284.

in the sky, not the idea. When I am thinking of pain, I am thinking of the bodily sensation, not of the idea of the bodily sensation. In either case, (1) would rule that the idea is not an idea!

The Cartesian definition is implied by the conceptualist solution to the traditional problem of universals. Philosophers from Plato onward have observed that all individual men have in common the property of being a man, or equivalently, that all are instances of mankind. Properties and kinds are therefore called universals, and the metaphysical questions concern their existence and nature. The conceptualist maintains that universals are concepts, and therefore exist in the mind. Thus when reflecting on the signification of general words like *horse* and *white*, Locke argues that "the abstract idea for which the name stands, and the essence of the species, is one and the same" (1690: §3.3.12).[4] Conceptualism can seem very natural, given that we use the devices of quotation and italics to refer to both the properties and the concepts expressed by general terms. Thus it might well appear that we are referring to the same thing when we refer to the property *white* and the concept *white*. But we are not. The property of being white is possessed by pieces of chalk and paper – by anything that is white. The concept of being white is not possessed by pieces of chalk or paper, since they cannot think. To possess the concept of being white is not to be a white thing, but to be able to think of white things.[5]

Since the Sun and the property of being a star are not mental events, not things that occur to us or things that we can have, and not things that we can *express*, they are not the sorts of thing the expression theorist should say determines meaning.[6] Using the Cartesian definition, in fact,

4 Compare and contrast Berkeley 1710; Husserl 1900: Chapters 1.4 and 2.1; and Horwich 1998a: 4, 21.

5 Since it is absurd to claim that properties and kinds are concepts, a more defensible form of Conceptualism maintains that there are no such things as properties and kinds: in addition to the particular horses and the concept of a horse there is no third entity that all the particular horses have in common. We sometimes find distinguished philosophers failing to note that the reductive and eliminative forms of Conceptualism are distinct and incompatible. See, e.g., Butchvarov 1996.

6 Locke elsewhere restricted "idea" to *immediate* objects of thought (see, e.g., Locke's second letter to Stillingfleet, quoted in A. C. Fraser 1959: §2), as did Leibniz (1709: §2.1.1). But "immediate" does not make sense in application to thought the way it does in application to knowledge or perception (cf. Reid 1785: 161–3). In particular, thinking of an object is not a causal process initiated by the object of thought, and we do not think of things in virtue of inference. In yet another place, Locke defined ideas as "whatsoever the mind perceives in itself, or is the immediate object of perception" (Yolton 1956: 91), as does Descartes (1641b: 68). By this definition, an itch is an idea, but an unconscious thought is not.

an ideational theory of meaning would seem tantamount to the referential (see §22.3), with the twist that referents are understood in light of an idealistic metaphysics.

A more fundamental problem is that unlike "idea," the phrase "object of thought" is not a genuine predicate-forming term when it means an *intentional* object (§6.3). We cannot coherently talk about the domain of "Φ" in "S is thinking of Φ," nor in "Φ is an object of thought" or "Φ is the object of T." That is, we cannot take "Φ" to be a quantificational variable when it signifies an opaque, intentional context (§6.4). If we try to, we get Meinongian absurdities. A woman who is thinking "That is the morning star" but not "That is the evening star" is thinking about the morning star but not the evening star. If "Φ is an object of thought" were a genuine predicate, it would follow that Venus both is and is not an object of thought, a contradiction. Or suppose that I am thinking about Santa Claus. While it is true to say "The object of my thought is Santa Claus," we cannot treat this as an identity statement, because Santa Claus does not exist. "There are objects of thought that do not exist" sounds contradictory, but is not. For it does not mean "$\exists x[Ox \,\&\, -\exists y(y = x)]$." All it means is that we can think of things that do not exist. Since "S is thinking of Φ" is not a relational predicate, thinking of Φ does not entail the existence of Φ. If we adopted the Cartesian definition, and took it seriously, we would have to conclude that there are ideas that do not exist. Descartes (1641b: 10, 52), and later Meinong (1904; 1910: 62), were led to distinguish two types of existence: the sort of existence Santa Claus and the Sun both have, namely existence as objects of thought, and the type of existence that only the Sun has, namely existence as a real object. The problem with this maneuver is that the idea of Santa Claus has the sort of existence that Santa Claus lacks. The idea really exists, Santa Claus does not. The idea of Santa Claus does exist "in the mind" in the sense that it occurs as a part of mind – that is, it is a mental event. But this does not mean that the idea is a figment of the imagination. The idea is not a fictitious or mythical being – not a *mere* object of thought.

Descartes did not intend his definition to have such consequences. His exact words were: *"an idea is the thing thought of itself, in so far as it is objectively in the understanding"* (1641b:9). The ensuing discussion makes it clear that the objective understanding clause was supposed to preclude the problems that we have discussed. But the clause is obscure, and it is not evident how it avoids the problems. Arnauld evidently grappled with the same difficulties, and officially defined an idea as *"anything in the*

mind when we think of a thing."[7] This is either equivalent to the Cartesian definition, or counts every mental event accompanying thought as an idea, depending on how "in the mind" is taken. In his "Introduction," Arnauld says that *"[t]he form by which we represent a thing to ourselves is called an idea."*[8] While more obscure, this formulation would seem to exclude the mere accompaniments of thought. But it would also seem to count words themselves as ideas, along with drawings.

According to Dickoff and James (1964: xxxviii), what Arnauld meant was that "*the idea is the thing as presented to the mind*." This is very similar to Descartes's obscure formulation, but has an additional defect. For example, it would seem that the Sun is presented to the mind as an external object in the sky – that is what we take the Sun to be. Hence it seems that we must say that the Sun as presented to the mind both is and is not an idea. Arnauld's Cartesian formulation ushers in all of the horrors of the sense-data theory, the view that some mental entity actually has all of the properties that physical objects appear to have.[9] Arnauld may have meant, though, that the idea is the presentation of a thing rather than the thing presented. His definition in a later work (Arnauld 1683: Chapter 5) supports this interpretation, and was cited approvingly by Reid (1785: 206).

(2) An idea is the presentation of a thing in thought.

We might call this the *Fregean definition* of "idea." Frege himself used "idea" to denote images. But "presentation" played the role in his theory of sense that "idea" does in standard formulations.[10] The Fregean definition suffers from the fact that "presentation" is a technical term that is as much in need of definition as "idea." For example, it would be very difficult to decide, using (2), whether sensory images count as ideas, or

7 Arnauld 1662: 33. See also Pappas 1995: 355: "whatever is immediately before the mind when it thinks."

8 This is similar to Descartes's obscure Definition 2: "*Idea* is a word by which I understand the form of any thought, that form by the immediate awareness of which I am conscious of that said thought" (1641b: 52). See also Descartes's (1641b: 105) reply to Arnauld. J. S. Mill's (1843: 4.2.1; 1865: Chapter 17) definition of an idea or concept was very similar.

9 See Broad 1923 and Barnes 1944–5.

10 See Frege 1892b: 57; 1918: 12–13; 1979: 131. Cf. Brentano 1874: 78–9, 198; Husserl 1900: §2.2.10; Marty 1908 as expounded in Liedtke 1990: 47; Meinong 1910: xii; Evans 1982: 62, 104; 1985; Peacocke 1981: 191; 1983: 89, 106; 1986: 15, 63; Fitch 1993: 471; and Fodor 1998a: 15–22. Arnauld himself uses "perception" rather than "presentation"; interpretation is difficult because Arnauld equates perception, knowledge, thought, and conception, and takes them to be one kind of representation (1683: 65–7).

conceptions. Reid was clearer, saying that the idea of Φ is the thinking (or conceiving) of Φ.[11] This is only slightly inaccurate: conceiving of Φ is the *occurrence* of the idea of Φ rather than the idea itself (Definitions 15.4, 15.5). As Reid's formulation highlights, however, ideas themselves can be said to have objects. The object of the idea of gold is gold, which is what we are thinking of when we are thinking of gold. It is the *objects* of ideas that can be said to be objects of thought, not the ideas themselves.

In §15.6 we discussed the fact that concepts are said to have *contents* as well as objects. We say that the Sun is the object of the concept of the Sun, and that "the Sun" is its content. The content of the thought that the sky is blue is *that the sky is blue*. We use either italics or quotation marks when referring to contents. The intentional object of the thought is the sky, and its relational object is the proposition that the sky is blue. Cartesians made the mistake of taking ideas to *be* the objects that they *have*. While less egregious, a similar mistake is made by some contemporary Fregeans who take concepts to be the *contents* that they have.

(3) An idea is a thought content or part thereof.

Thus Peacocke says, "I use 'content' and 'Thought' as stylistic variants" (1986: 1). And Boghossian says that "concepts just are what thought contents are composed of" (1998b: 257).[12] On this *neo-Fregean definition*, the thought that the sky is blue is the content, *that the sky is blue*. Part of that content is the content, "the sky," which is identified with the concept of the sky. I do not believe that such claims make much sense. First, neo-Fregeans presuppose that in *oratio obliqua* expressions of the form "the thought that p," the clause "that p" is functioning as a singular term. This seems to be contrary to the linguistic evidence (cf. §21.2). Second, "the content that the sky is blue" does not have the same meaning that "the thought that the sky is blue" has, at least if we confine ourselves to conventional English. Indeed, the expressions do not even seem to have the same syntax: to make sense of the former, I have to hear a comma pause after "content" or, equivalently, to hear "the sky is blue" as italicized or quoted, making it analogous to "the mayor, John Doe."

There are material as well as linguistic differences between thoughts and contents. The thought that the sky is blue has the content "the sky is blue"; but it makes no sense to say that either the thought or the content

11 Cf. Ockham, *Summa Logicae I*: §12.

12 See also Husserl 1900: §1.4.30; Burge 1979b: 537; Peacocke 1986: 14–15; 1992: 203; and Schiffer 1982: 137; 1987a: 14–17.

has the *thought* "the sky is blue." Thoughts occur to people, and have certain causes and effects. People have thoughts and acquire concepts. Thoughts and concepts have contents. It makes no sense to say such things of contents. Contents are individuating properties of thoughts; thoughts are not properties of thoughts. The neo-Fregean definition is thus a subtle category mistake.[13] The property of having the content μ should be identified not with the thought μ, but with the property of being identical to that thought or having it as relational object.

The Cartesian definition of an idea and its neo-Fregean variant are nothing but trouble if taken literally and seriously. The fact that Cartesians offered many different definitions seemingly without awareness of their distinctness compounded the problems, which undoubtedly led to much of the disrepute into which ideational theories have fallen. I prefer to view Descartes, Arnauld, Locke, and Leibniz as innovators who grasped some important truths unclearly, and seek to present a clear and defensible theory of ideas that preserves their important insights about meaning.

§19.2 IDEAS VERSUS IMAGES

Many scholars, particularly the sensationist philosophers and psychologists, have defined "ideas" as *sensory images.*[14] Whereas modern sensationists defined the idea of Φ as the inner-speech word "Φ" (see §19.3), the classical sensationists defined the idea of Φ as the image of Φ. One identifies the idea of an elephant with the image of an elephant, which might be visual, auditory, or tactual. The other identifies the idea with the inner-speech image of the word "elephant," which is necessarily auditory (at least in hearing speakers). Treated as stipulations, there can be little objection to either sensationist definition. But the sensationists went on to assign to what they called "ideas" a theoretical role that can be filled only by thoughts and their parts. The associationists held that to think of something is to have an image, and that to mean something is to

13 The neo-Fregean definition of concepts as contents does fit naturally with the Fregean identification of meanings with concepts. But we will criticize that version of the ideational theory of meaning in §21.1.

14 See Descartes 1641a: 159–60 (but contrast 185–6); Berkeley 1710: Introduction; Hume 1739: §1.1.1; Condillac 1746: §16; Hartley 1749: ii, 76; J. Mill 1829: Chapter 4; Frege 1884: §27, n. 1; §60; 1892b: 59; Bain 1855: 315; Titchener 1897: §43, §83; 1909; 1914; Wundt 1911: 45; Warren 1921: 297–9; Hampshire 1939–40; Goodman 1952: 67. See also Hobbes: 1651: Chapters 1.2 and 1.4; 1655: Chapter 2 (discussed in Hungerland & Vick 1981: 75, 89); Peters 1956: 134–5; Hacking 1975: 23–4; Bickerton 1981: 221–2; Carruthers 1992: 5–6.

express an image. In most cases, the overriding premise was the sensationist thesis that sensations and images are the only introspectible mental states. These substantive claims are untenable. To strengthen our understanding of thought-parts, and to preclude possible confusion, we will set out the distinction between images and thought-parts in detail. On our definition, "idea" and "concept" are equivalent and denote cognitive thought-parts. So in our terms, this section and the following will compare and contrast ideas or concepts with images.

We will look first at the relationship between *the concept of Φ* and *an image of Φ*. There are many similarities. Both the idea of a horse and an image of a horse, for example, are mental events that occur to us from time to time. Both in some way represent horses. Both are typically parts of larger wholes: the idea of a horse might occur as part of the thought that a horse is jumping over a fence, and an image of a horse might occur as part of an image of a horse jumping over a fence. Having the concept of a horse, moreover, is highly correlated with the ability to form an image of a horse. Both images and concepts have intentional objects. I can think of Pegasus, and form an image of him, even though Pegasus does not exist. Both images and ideas, finally, are introspectible mental events. I can tell introspectively whether I am thinking of a horse as surely as I can tell whether I am forming an image of a horse.

Despite these similarities, *concepts or ideas are not sensory images*. The fundamental difference is that images, unlike ideas, are composed of sensations. As a result, images are localizable and particularized in ways that ideas are not. Moreover, people have concepts of things that they do not, and sometimes cannot, form images of. Finally, concepts and images differ in the ways in which they correspond to reality, or fail to do so.[15]

Let us take the most obvious difference first: *People have concepts of many things that they cannot form images of*. Consider the concept of a thousand-sided figure, Descartes's (1641a: 185–6) example. I can think about thousand-sided figures (as I am doing now); I know many things

15 For an excellent discussion of the relationship between images and thoughts, with reference to the empirical literature, see Fodor 1975: 174–95. See also James 1890: Chapter 18; Ryle 1951; 1958; Ginnane 1960: 387; Aune 1967a: 102; Blackburn 1984: 47; and Stillings et al. 1987: 36–48, 449–50. The distinction was very clear to Arnauld (1662: Chapter 1) and Frege (1884: §27, n. 1; §60; 1892b: 59; 1894: 79), although Arnauld used the term "idea" generically to cover both, and Frege restricted it to images. Reid (1784: 398, 426–7, 517) realized that we could conceive of things we cannot imagine, but wrongly concluded that images are just a species of conception, and that there are no such things as images as objects of thought. Humphrey (1951: Chapters 2–4) summarizes the "imageless thought" controversy in introspective psychology.

about thousand-sided figures (such as that they have two hundred times as many sides as a pentagon); and if I put my mind to it, I could even prove theorems about such figures (such as what the sum of their angles is). But I cannot form an image of a thousand-sided figure. The image would have to be visual or tactual, but in neither case can I form a distinct image with that many sides at once. My "image screen," we might say, simply does not have enough "resolution." Consider a more radical example. I have the concept of the number 2. But I cannot form an image of the number 2 any more than I could draw a picture of it. I can easily enough form images of the numeral "2" or of the word "two," or even of two sticks. But numerals, words, and sticks are not numbers. Numbers are completely abstract, and cannot be seen, heard, felt, or otherwise perceived. Consequently I have no idea how to form an image of a number, and no image I can form would count as an image of a number. Finally, it is plausible that the congenitally blind cannot form visual images. Hence they cannot form any image of shadows or rainbows, for example. Nevertheless, a blind woman may have the concept of a rainbow. Indeed, she could be the world's leading expert on the physics of rainbows.

Second, even when the subject has the ability to form an image, *the concept or idea of something can occur to S when an image of it is not occurring*. When I read a novel, for example, I sometimes take the time to form a sensory image of the scene the author is describing. But usually I do not. When I am reading out loud, my attention is focused on the words on the page; it would thus be very hard for me to form a visual image of a horse as I read the word "horse," even though I am certainly thinking about horses. Now it might be insisted that in cases like these, I really am forming an image, but too briefly for introspection to detect. While this is admittedly a possibility, there is no reason to think that it is true. On the contrary, the fact that people can think of chiliagons and numbers without forming images of them, and that the congenitally blind are able to think about rainbows without being able to visualize them, is a reason to believe that I can think of horses without forming an image of horses. It might more plausibly be suggested that when thinking about horses I must be forming *some* related image.[16] Thus when I am thinking of horses, I do normally say the word "horses" (or some relative) to myself, which means that I am forming an auditory image of the word "horses." But this suggestion does not contradict the point immediately in question, which is that a

16 See, e.g., Stout 1899: 394 and Titchener 1909. Contrast Humphrey 1951.

person can think about horses without forming an image of horses. For an image of the word "horses" is not an image of horses.[17] Nor does the suggestion contradict the more important claim that a thought cannot be identified with any image that might accompany it.

Third, and conversely, *an image of something may occur to a subject when the concept of it is not occurring*. This occurs whenever the subject has an image of something that he does not recognize or incorrectly identifies. A young girl fascinated by the peculiar bird that she sees at the zoo (an emu) may later form images of it without knowing what kind of bird it is, and without knowing the difference between an emu and an ostrich. If she forms a memory image of the bird she saw, she will have an image of a ratite even though she does not have the concept of a ratite and is not thinking of ratites. Similarly, an image of an ambiguous figure may be equally naturally accompanied by two different thoughts. Thus a man may have an image of a sunset without thinking of a sunset because he is thinking of a sunrise (or vice versa).

The first three facts cited show that ideas and images are not perfectly correlated. Their lack of correlation is a consequence of the fact that the two sorts of entities differ markedly in their intrinsic nature. *Images are composed of sensations, and are very much like sense impressions*. Consequently every image is *locatable* in one of our sensory fields, and is either visual, auditory, tactual, olfactory, or gustatory. Moreover, images differ from person to person and change from second to second. Unlike sense impressions, images can be faint or vivid, and are subject to voluntary control. I am forming an image of a horse now. It is a visual image of a brown horse, with a black tail. Hence the image is a two-dimensional arrangement of sensations of brown and black, and is in many ways analogous to an impressionistic painting of a brown and black horse. The horse is right in the middle of my computer screen. That is, my image of the horse is located in the middle of my sense impression of the computer screen, both of which are located in the center of my visual field. The image is rather faint: I cannot make out individual hairs in the tail. I can change the details of my image at will: I just made the horse's tail green, and turned the horse upside down. The location of an image can also

17 In his rejection of ideas, Ryle asserted that "[i]f he were told to think the thought of 'Lillibullero,' without producing, imaging, or actually listening to the tune itself, he would say that there was nothing left to think" (1949: 227). But we can think the thought that Lillibullero was a tune mentioned by Ryle without singing the tune in our heads or out loud. Thinking of Lillibullero may even have, as one of its effects, a futile and frustrating attempt to recall how the tune goes.

be changed at will: I have now moved my image of the horse above the screen, and now below it.

Concepts are not composed of sensations. The concept of a horse cannot be described as faint or vivid, nor as visual or auditory, nor as above or below my sense impression of the computer screen. It is almost completely unlike a picture of a horse. I cannot change the concept of a horse at all. The concept of a brown horse does have at least two components: the concept of brown and the concept of a horse. But the syntactic relation between the components is quite different from the spatial relations among the sensations in an image.

Since images are composed of sensations, they are *particularized* in ways that concepts are not. Thus any image of a horse is an image of a brown horse, or of a white horse, or of some other color horse. But the concept of a horse is not the concept of a brown horse, or the concept of a white horse, or the concept of any other color horse. The image of a horse that I form at one time will normally differ in detail from the image of a horse that I form at another time, and from the image of a horse that anyone else has formed. There is only one concept of a horse, however, which has occurred to me at various times, and to other people as well. The concept of a horse is an abstract, unchanging entity, a fixed type of event. People's conceptions of horses may change over time, and the word "horse" may come to express different concepts. But that does not mean that the concept of a horse changes. The concept of a horse occurs to someone when, and only when, the person is thinking about horses.

Images are not only particularized, they are *particulars*. That is, the general term "image," along with more specific terms like "image of a horse," denote event-*tokens*. The extension of such terms is a class of dated occurrences. There is no such thing as *the* image of a horse, *simpliciter*. There is only the image formed by John at 10:00, the image formed by Mary at 10:01, and so on. No matter how similar John's and Mary's images are, there are two images. There is of course a type whose tokens are all of the images of a horse. But this type is not the referent of "the image of a horse" in conventional English. Concepts, by contrast, are *universals*. The term "concept" applies to event-*types*. The concept of a horse conceived by you right now is not only similar to but identical to the concept conceived by me right now. It is for this reason that images, like sensations in general, cannot exist apart from the particular individuals who have them, while concepts can. Note that whereas "*a* concept" is a well-formed general term, "*a* concept of a horse" is not. Since the concept of a horse, the concept of a cat, and the concept of a ship are

all concepts, "a concept" applies to all of them. But there is only one concept of a horse.[18] This concept has many occurrences, of course, but "a concept of a horse" does not apply to them in any standard sense. "Thought," "belief," "desire," and other terms for propositional attitudes have a similar "logic." For example, "*A* thought" is a well-formed general term, applying to countless event-types, such as the thought that the sky is blue, the thought that grass is green, and so on. It also applies to all of the different occurrences of the thought that the sky is blue. But "*a* thought that the sky is blue" is not well formed (at least in standard English) – there is only one thought that the sky is blue. In particular, the many tokens of this event-type are not conventionally referred to as "thoughts that the sky is blue."[19]

Our memory for images is generally not accurate enough for us to be sure that an image that we are forming now is exactly the same as one that we had before. I am confident that the image of Pegasus that I am forming now is nearly the same in whiteness as the image that I formed yesterday, but I cannot be absolutely certain. By contrast, to remember that I was thinking of Pegasus yesterday is to know that the very same idea – that of Pegasus – was occurring to me yesterday. If a different idea was occurring to me, then I was not thinking of Pegasus.[20]

Note that the following contexts have different logical properties.

(4) S is forming an image of Φ.
(5) The idea of Φ is occurring to S.

Both contexts are intentional: the objects of both images and concepts may be nonexistent, as we observed. Nevertheless, (4) but not (5) is subject to the substitutivity of identity. If S is forming an image of John F. Kennedy, then S is forming an image of the thirty-fifth president, whether he knows it or not. By contrast, S may be thinking of John F. Kennedy even though the idea of the thirty-fifth president is not occurring to him. Indeed, it may never have occurred to him that Kennedy was the thirty-fifth president.[21] Similarly, an image of Kennedy is an image of a World War II veteran. Yet

18 Contrast J. S. Mill 1869: 237. Note the seeming absurdity of Linsky's claim that "*the sense of 'red' in English* and *the sense of 'rosso' in Italian* denote the same sense, and each expresses a different concept of it; hence they express different concepts of the same concept" (1983: 55).

19 See note 11 to §12.3.

20 Compare and contrast James 1890: Chapter 9; pp. 231–7.

21 In the terms used in §6.2, "of Φ" is a *transparent* description of the object of S's image in (4), an *opaque* description of the object of the idea in (5). Since images are not conceptual, it is hard even to think up an opaque sense for (4).

the idea of Kennedy may occur to someone without the idea of a World War II veteran occurring to him.

Finally, *images and concepts differ in the ways in which they correspond to reality*. An image of the Eiffel Tower, for example, may be described as "lifelike," "true-to-life," "accurate," perhaps even as "veridical"; negatively, images can be "distorted," "upside down," and "too thin." None of these adjectives can be applied to the concept of the Eiffel Tower, even with stretching. On the other hand, images do not "apply to" objects in the way that concepts do: the concept of a horse applies to an object iff the object is a horse; it makes no sense to say that an image of a horse applies to an object. Similarly, images cannot be assessed as true or false in the way that some ideas can be. For there is nothing in the phenomenon of imagery corresponding to the difference between propositional and nonpropositional ideas. Consider the difference, for example, between the idea of *the tall man in the corner* and the idea that *the man in the corner is tall*. The idea that the man in the corner is tall may be true or false depending on whether or not it corresponds to reality – depending, that is, on whether the man in the corner is in fact tall. The other properties of the man are irrelevant to the truth or falsity of that idea. The idea of the tall man in the corner, on the other hand, is not the sort of thing that can correspond or fail to correspond to reality. Any image of the tall man in the corner, by contrast, will also be an image of the man in the corner's being tall. And whether an image of the tall man in the corner is true-to-life depends on more than that man's being tall and in the corner. It would be very unlifelike, for example, if it was also an image of a man with an elephant's trunk. In general, images resemble pictures and iconic representations, while ideas resemble words and symbolic representations.

The difference between ideas and images parallels that between *imagining* and *forming an image*. The latter involves sensory images, which are composed of sensations. The former involves thinking thoughts, which are composed of ideas. To *imagine that p* is at least to think the thought that p without believing that p, and therefore without remembering, perceiving, or inferring that p. If you ask me to imagine that I am lying on the beach, I would undoubtedly form an image of the beach. But if you ask me to imagine that the demand for wheat is inelastic, I would not know how to form an image of inelastic demand. I might form an image of a graph or equation representing an inelastic demand, but then again I might not.

The facts presented in §19.2 show that the concept of Φ cannot be identified with an image of Φ, and that thinking of Φ is far from perfectly correlated with having an image of Φ. It does not follow that thinking does not require images in some way, though. Given all that has been said so far, it remains possible that thinking of a thousand-sided figure requires some image *other* than the image of a thousand-sided figure. A natural candidate is provided by *inner speech*. We can summarize the preceding section by saying that the concept of Φ cannot be identified with the image of Φ because the image is a picturelike representation, while the concept is wordlike. But images of *words* are a special case, being both picturelike and wordlike. The inner speech word "dog" is an image not of a dog, but of the spoken word "dog." Inner speech consists of a temporal series of verbal images of words in the speaker's own voice organized into a discourse.

All of the similarities between concepts and words noted in §13.1 and §14.1 hold for inner-speech words. Indeed, inner-speech words are more like concepts in many respects. Both are mental representations. Both thinking and inner speech are structured, and both are subject to voluntary control. Both are introspectible events that are not publicly observable, in marked contrast to external speech. Both are neutral with regard to belief and desire. That is, a subject who says to himself "It is raining" may or may not believe that it is raining, and may or may not want it to rain. Finally, there is at least a strong connection between thought and inner speech. The thought that it will rain typically occurs to us when the sentence "It is raining" (or something equivalent) occurs in our inner speech. Conversely, we typically say to ourselves "It is raining" when we think "It is raining." So the modern sensationalist thesis that the concept "Φ" is the inner-speech word "Φ" is plausible, and avoids many of the objections to the classical sensationalist thesis.[22]

22 Cf. Plato, *Theaetetus*: 189e-190a; Condillac 1746: 2.401b; Titchener 1914: Chapter 11; Warren 1921: 148–9; Wittgenstein 1953: Part I, #329; Sellars 1958: 218; Devitt & Sterelny 1987: 117; Goschke & Koppelberg 1991: 156. Watson (e.g., 1930: Chapter 10) transformed the thesis behavioristically, maintaining that thinking is subvocal speech (movements). Contrast Humphrey 1951: Chapter 8; Price 1953: 235ff.; Geach 1957a: 23, 99; Aune 1967b; Vendler 1977; and Bach 1987a: 63, fn. 11. Ryle (1951: 258–9, 264, 267; 1958: 405–6; 1979: Chapter 5) tried to have it both ways, denying that thinking could be identified with inner speech while insisting that it was not something distinct from inner speech, either. James (1890: Chapter 9, pp. 278–83) was similarly enigmatic. He seemed to clearly

Despite these similarities, the thesis that thought is identical to inner speech is thoroughly untenable. First, inner speech can occur without thought, as when a woman utters a meaningless sentence to herself, or a sentence in a language that she does not understand. This raises the question, "What distinguishes thoughtful from thoughtless inner speech?" The obvious answer – that one but not the other is accompanied by thought – is not available to the sensationalist. For it entails that the thoughts are distinct from the inner speech that they accompany. The sensationalist would have to say something like this: inner speech is thought if and only if it is meaningful. But what would make inner speech meaningful other than the fact that it is expressing a thought? It is not enough that the words occurring in inner speech have a meaning in some language. For even if S happens to know that language, there is no guarantee that S was using it on that occasion. And even if he was, nearly all words are ambiguous. It is not the mere fact that "bank" is occurring in S's inner speech that makes it true that S is thinking of a commercial bank rather than a riverbank. The account of speaker meaning given in Part I is not open to the sensationalist, for that account maintains that what a speaker means by a word is determined by what the speaker intends, and therefore by what he thinks. The sensationalist would get nowhere by claiming that the meaning of one piece of inner speech is always determined by the meaning of another. A sensationalist would have to explain the difference between thoughtful and thoughtless inner speech, but does not have the resources to do so.

There can also be thinking without inner speech, as when a subject uses visual images. For example, I can think about a complex dive that I do not have the words to describe, by visualizing it. And a person who is trying to fit a piece into a puzzle by trial and error is thinking all the time about where the piece goes, but does not have to be continuously saying to herself "Where does it go?" There is abundant evidence that children think before they have learned a language, and that other animals think even though they are incapable of learning a language (see, e.g., Weiskrantz 1997). In the tip-of-the-tongue phenomenon, we think of a person without being able to recall the name. And there is evidence that certain forms of aphasia garble inner speech without affecting thought

distinguish the idea from the words, but nevertheless says they are "consubstantial." Stout (1899: 395) was clearer, distinguishing ideas from inner speech words, and defining an idea as the combination of an image and a "meaning"; cf. James 1890: Chapter 18, p. 49; Harman 1968: 69.

(Humphrey 1951: 251). So the correlation between thought and inner speech is far from perfect.

Most critically, the facts distinguishing thoughts from sentences and words from concepts (§13.1) also prove that ideation is distinct from inner speech even when we think to ourselves in words. Two subjects who say "Every man is an animal" and "All men are animals" have different inner speech even though they are thinking the same thought. And two subjects who utter the sentence "Flying planes can be dangerous" to themselves have the same inner speech, but may be thinking very different thoughts. A man may be thinking a different thought each time he says to himself "She is beautiful." What the subject is thinking in those words is the occurrent factor that *determines* what the inner speech means and refers to. The inner-speech words have their meaning and reference contingently; the ideas that they express have their content essentially.

Moreover, our thoughts are almost always more complex in some respects than our inner speech. For example, when Steve says to himself "Triangles are three-sided polygons," he could well be thinking the thought that *all and only* triangles are three-sided polygons, even though there is no element of the sentence corresponding to "all and only." In the "flying planes" example, the elements of the thought expressed by "flying" and "planes" may be related as they are when we think "Planes are flying," or as they are when we think "They are flying planes"; but the inner-speech words just occur one after the other in either case. In other respects, inner speech appears to be more complex than thought. For inner speech can be analyzed into a sequence of speech sound images that have no known correlates in the case of thought. An image of the "F" sound is part of the occurrence of "Flying planes can be dangerous" in inner speech, but does not represent any part of a thought.

Other differences concern the temporal nature of thought and inner speech. *Inner speech is an episodic, serial process*. First, to say "It is raining," either to ourselves or out loud, we first have to say "it," then "is," then "raining." We have not said "It is raining" until the whole sequence has occurred. It thus *takes time* to say "It is raining," usually on the order of a second. We can speed up or slow down the process at will. Second, we say one thing after another to ourselves, but never say two things at once. In this respect, inner speech is like external speech. By contrast, *thinking appears to be a continuous, parallel process.* The components of a thought do not appear to occur to us one after the other. That is, thoughts do not seem to "take time." Thoughts may occur *for* a significant period of time. But at each moment of the interval it will be true that the thought has

occurred.[23] Thus the thought that it is raining occurs continuously during the time that it takes us to say "It is raining" to ourselves. Furthermore, we seem capable of thinking many different thoughts simultaneously. While I am writing this sentence, for example, it is occurring to me to strike the keys; it is simultaneously occurring to me that words are being printed on the screen, that my son is making noise, that the light is on, and so on. Of course, it is possible that thought too is a serial process with rapid switching. That is, it may be that all of the different thoughts just mentioned occur to me in sequence, but in such a short time that they seem to be simultaneous. Similarly, a thought may actually take time to occur, but too little time for us to detect the stages of the process using introspection. But whether thought actually is, or only simulates, a parallel, continuous process, it differs markedly from inner speech.

Finally, some of the facts distinguishing the idea of Φ from the image of Φ (§19.2) also serve to distinguish ideation from inner speech. Inner speech is composed of auditory sensations, and thus has a location in our auditory field (specifically, in the center). Thoughts are not composed of sensations, and have no location in the auditory field. Inner speech has a definite pitch, timbre, tempo, and rhythm, which differ from person to person and from time to time. The same adjectives cannot be applied to thoughts.

One piece of introspective evidence supports the identification of inner speech and thought, at least when the two occur together. It may lie behind the persistent attractiveness of the identification in the face of heavy counterevidence. Ryle might have put it this way.

> When we are thinking to ourselves (or thinking out loud, for that matter), there do not seem to be two things going on. There just seems to be one process. The relationship between the thoughts I am thinking and my inner speech is distinctly different from the relationship between my inner speech and the music I am hearing in my head at the same time or the sequence of sense-impressions I am receiving. The latter appear as separate processes, one sometimes interrupting the other.

I do not wish to deny these appearances. I believe that they indicate that there is an especially close relationship between inner speech and thought when we are thinking to ourselves. The evidence cannot show, however, that inner speech and thought are identical. For we have reviewed

23 In a mystifying non sequitur, Ryle (1962: 447–8) concluded from this fact that there is no such thing as conceiving concepts, and that "operating with a concept" is not an "incident."

overwhelming evidence, much of it introspective, that they cannot be identical. So there must be some other explanation of the single-process appearance. There often is. For example, we do not normally perceive a concert pianist as doing two things at once, playing the left-hand part and playing the right-hand part. We normally just hear an integrated performance. But of course, the pianist is doing two different things at the same time. Taking the analogy a little deeper, with considerable practice a pianist can play scales one octave apart with two hands in such a way that the listener simply hears one scale with a special color. It is also possible to play the two scales so that they sound completely separate. Some Bach specialists make his four-part fugues sound as if they were being played on four separate pianos. Others make them sound like one line with a complex sonority. We often perceive distinct sequences of simultaneous events as a unit when they are related in certain ways. The conclusion to draw, therefore, is that inner speech and the thought accompanying it are perceived as a single process because of the way in which the two processes are integrated, not because the inner speech and the thought are identical. The ability to think in a foreign language, acquired after extensive practice, is the ability to integrate thought with speech in this way.

I am not claiming that, instead of thinking in words or pictures, we really think in ideas. "Thinking *in* ideas" makes little sense. The occurrence of ideas *is* thought, not a vehicle or accompaniment of thought. When we think in French, the French sentences that we produce are not themselves our thoughts, and our uttering the sentences is not our thinking, although the sentences and utterances are intimately related to the thoughts they accompany.[24] That relationship differs markedly from the relation between the thought and the ideas composing it. The fact that ideation differs from imagery therefore does not preclude important connections between the two phenomena. Indeed, nothing that we have said refutes the very general Thomistic thesis "that the soul never thinks without an image" (*Treatise on Man*: Q 84 Art. 7). All that we have established is that thinking and ideation do not consist of images, and that the idea of Φ is not tied to an image either of Φ or of the word "Φ." Whether or not an occurrence of the concept of Φ requires the occurrence of some image or other is not an issue that we need to decide.

24 Cf. Husserl 1900: 282 and Vendler 1977: 60. Contrast Wittgenstein 1953: §329–32; Price 1953: 300; Ryle 1951: 258–9; 1958: 394–9, 406; Ginnane 1960: 376; Aune 1967b; Sellars 1969: 104–13; 1979, Chapters 4–5; Vendler 1972: 36–7; Chastain 1975: 235; Devitt & Sterelny 1987: 117; Carruthers 1989: 100–1; Gaulker 1994: e.g., 26.

We observed in §12.1 that while "thinking the thought that p" denotes the simple occurrence of the proposition that p, "thinks that p" means believing that p. The verb "conceive" has the same ambiguity. It can be used to express the act of conceiving subpropositional concepts, an act that implies neither belief nor disbelief, as in "S is conceiving of Φ" and "S is conceiving Φ" (§15.3). "Conceive" can also be used to express propositional thought, as in (6), which can mean "S is thinking of Φ as Ψ."

(6) S is conceiving of Φ as Ψ.

Thus while reading *Animal Farm,* I was conceiving of pigs as being more intelligent than people; and when I watch *Psycho*, I end up conceiving of Tony Perkins as a murderer. It does not follow that I have ever believed pigs to be intelligent, or Perkins to be a murderer.

But "conceive" can also be used to express belief. To say that, unlike Ptolemy, Copernicus conceived of the Earth as moving is normally to say that Copernicus believed that the Earth moves.

(7) S conceives of Φ as Ψ.
(8) S conceives Φ to be Ψ.
(9) On S's conception (concept), Φ is Ψ.
(10) S has a conception (concept) of Φ.

Formulas (7), (8), and (9) entail that S believes that Φ is Ψ. Woodfield (1991: 552) observed that the past tense, "conceived," is ambiguous between thought and mere belief; consider "When I was young, I used to conceive of spiders as crabs." When it expresses belief, (7) is not a frequentive, and does not stand to (6) as "plays" stands to "is playing."

Similarly, the noun "conception" has both cognitive and cogitative senses. It can be used to designate the act of conceiving concepts, an act that implies neither belief nor disbelief. And as we are using the term "concept," it designates parts of propositions, which in many cases are not even capable of being believed. But "concept" and "conception" can also be used as in (9), to designate a belief system. Whereas having the concept of something does not entail having any beliefs, having a conception of it does: (10) means that S has some beliefs about Φ, and (9) specifies a particular belief that S has. The contrast may be brought out as follows. Both Ptolemy and Copernicus had the concept of Earth. Both could think about Earth, and did so on numerous occasions. In this respect, the two

men were alike. Since they thought about the same thing (Earth), they had the same concept. Nevertheless, Ptolemy's conception of earth differed radically from Copernicus's. For Ptolemy believed that Earth stood still, while Copernicus believed that Earth moved around the Sun. A particular conception of Earth may be described as "revolutionary," "profound," "detailed," "correct," "well-supported," "scientific," "religious," and so on. Such adjectives do not apply to the concept of Earth.

Conceptions

The cognitive sense of "conception" may be defined with sufficient precision and accuracy for our purposes as a *centered belief system*, a system of beliefs with a common subject. *S's conception of Φ consists of S's beliefs about Φ*. "Concept," "idea," and "notion" all have the same meaning in conventional English when preceded by a possessive. When preceded by the definite article, these terms all denote thought-parts. Thus *Ptolemy's* concept of Earth and *the* concept of Earth are very different things. Ptolemy's concept of Earth consists of what he believed about Earth, and so contains the proposition that Earth stands still but not the proposition that Earth moves. The concept of Earth is not an object of belief at all, and is part of both the proposition that Earth stands still and the proposition that Earth moves.[25] *John's* idea of happiness may be unique (no one else has the same beliefs about happiness), even though *the* idea of happiness is ubiquitous (the idea crosses everyone's mind from time to time). And Hilary's concept of an elm tree may be pretty much the same as his concept of a beech tree ("a big, deciduous tree grown in the eastern U.S."), even though the concept of an elm is not the same as the concept of a beech, and even

25 The ambiguity of "concept" has generated a familiar argument for the "incommensurability" of theories after "paradigm shifts": (1) Ptolemy and Copernicus had radically different concepts of the Sun; (2) Ptolemy and Copernicus used "the Sun" to express different concepts; (3) Their theories about the Sun are about different objects; (4) Their theories are not really incompatible, and one cannot be described as an improvement on the other. The inference from (1) to (2) is sound only if "concept" means "belief system." The inference from (2) to (3) is valid, however, only if "concept" means "thought-part." So the argument is a grand fallacy of equivocation. Cf. Kuhn 1970: 149; Hamlyn 1971: 9–11; and Devitt & Sterelny 1987: 180. Kiel (1989: 15–23) makes the same mistake in a different way, presenting Fodor's argument that children could not communicate with adults if development involved qualitative changes in "concepts." So do Margolis and Laurence (1999: 48–9), who see a problem for the "theory theory" of "concepts" because subjects must have the same "concept" in order to disagree; the object of their criticism is represented in Murphy & Medin 1985: 448. Contrast Carey 1991: 465–8; Rips 1995: 77–9; Woodfield 1997: 86; Higginbotham 1998: 157; Millikan 1998b: 534–39, 544; Jacob 1998.

though Hilary has both concepts (if he did not have both, he could not know that elm trees are not beech trees, that only the former are properly called "elm trees," and so on.) In accordance with Definition 15.1, we will always use the terms "concept" and "idea" to mean *cognitive thought-part* (proper or improper), reserving "conception" for either a *centered belief system* or the act of *conceiving*.[26] Our goal in this section is to clarify the distinction between centered belief systems and thought-parts, so we will use the terms "conception" and "concept" accordingly.

To produce a better definition of a conception, we would have to address two problems. First, while it seems clear that S's conception of Φ contains *only* S's beliefs about Φ,[27] it is doubtful that S's conception contains *all* of S's beliefs about Φ. Suppose that Steve believes that Terry's tie has a speck of dust on it, but Bill does not. It is not clear that Steve's conception of Terry therefore differs from Bill's; the beliefs in this case seem too inconsequential. More significantly, the fact that John believes that Earth is called "Earth" while Jean believes that it is called "Terre" does not imply that their conceptions of Earth differ. S's conception of Φ excludes S's beliefs about how speakers or languages refer to Φ.

Second, in defining a conception as a centered belief system, meaning thereby a system of beliefs "about a common subject," the term "about" cannot be taken purely relationally (see §6.3). If S believes that Johnson rather than Kennedy was the thirty-fifth president, then S's conception of Kennedy differs from his conception of the thirty-fifth president, even though Kennedy was in fact the thirty-fifth president. On the other hand, "S's conception of Φ is identical to S's conception of Ψ" may be true even though the concept of Φ differs from the concept of Ψ. Thus my conception of Kennedy is identical to my conception of the thirty-fifth president. Whether S's conception of Φ is identical to S's conception

26 Woodfield (1991: 549; 1997: 83–4) defines the concept-conception distinction similarly, although he describes concepts as "classificatory norms." Katz's (1972: 450–2; 1977a) "dictionaries" and "encyclopedias" may be viewed as formal representations of concepts and conceptions, respectively. Perry (1990: 18–19) distinguishes between "notions" and "files," the latter consisting of a number of beliefs with a common notion. Higginbotham (1998) draws a related distinction between the concept of Φ and S's conception of *the concept*. S's conception of the concept of Φ consists of S's beliefs about the concept, such as that it applies to Φ.

27 Macià (1998: 183) suggests that conceptions may also contain emotional elements, since a person who has not been in war cannot have the same conception of war as someone who has. I think that is in large part because the lucky one does not know how horrible war is, for example. That is, without actual memories, the lucky person cannot have the demonstrative belief "That is how horrible war is" or "That's what war is like."

of Ψ is determined entirely by what S believes about Φ and Ψ. If S believes the same things about Φ and Ψ, then his conception of Φ and Ψ are the same even if the concepts are different. This leads to technical issues concerning the exact set of beliefs taken to define a conception and the precise interpretation of "identity" as applied to conceptions in any particular statement. If the set of beliefs defining a conception must have a common subject concept, then my conception of Kennedy will not be *numerically* identical to my conception of the thirty-fifth president. There will be two conceptions that are identical in the weaker sense of "exactly similar in relevant respects." The relevant respects would be the predicate concepts in the beliefs. If we wish to say that there is just one conception here, containing my belief that Kennedy was young as well as my belief that the thirty-fifth president was young, then "beliefs about a common subject" will have to be construed more broadly than "beliefs containing a common subject concept." These issues will not concern us. For no matter how they are resolved, conceptions differ fundamentally from concepts. It is the term "concept" that we need to clarify.

Differences

In order to be sure that there is no confusion between concepts and conceptions, and to deepen our understanding of thought-parts in the process, we will explore the differences between concepts and conceptions at length. First, there is no such thing as *the* conception of something. "Conception of Earth" is a general term whose denotation is an actually large and potentially unlimited set of centered belief systems. "Concept of Earth," on the other hand, necessarily denotes just one thought-part. It is thus as unnatural to think of "concept of Earth" as a general term as it is to think of "cube root of eight" as a general term. "Conception" never means just "part of a proposition." The term is ambiguous, as we have noted, but the ambiguity is different from that of "concept."

Being belief systems, conceptions of objects may be characterized as true or false, accurate or inaccurate, rational or irrational, fruitful or barren, complete or incomplete. None of these adjectives apply to concepts of objects. Lamarck's conception of evolution was false. Darwin's was much more accurate. The concept of evolution is neither true nor false: it is a part of some true propositions (e.g., the proposition that evolution occurs through natural selection) and of some false propositions (e.g., the proposition that evolution occurs through the inheritance of acquired characteristics). To say that the concept of Earth is a mental representation

of Earth is not to say that the concept is a "mirror" or "reflection" of Earth. Those metaphors might apply to conceptions, but not to concepts. The fact that the truth and completeness of a conception depends on the external world does not imply that either concepts or conceptions are so dependent. Note that "conception" inherits the act-object ambiguity of "belief." If we say that Einstein's conception of space was closer to the truth than Newton's (or vice versa), we are referring to the sets of propositions that Einstein and Newton believed about the Earth. That is, we are concerned with *what* they believed. If we say that Einstein's conception led him to predict that the apparent positions of certain stars should be altered when they are in the same direction from the Earth as the Sun, we are referring to Einstein's *believing* what he did.

Of particular importance to us is the fact that individual words are used to express concepts rather than conceptions. If Ptolemy and Copernicus had spoken English, they both would have used the word "Earth" to express the concept of Earth. They could not have used that word or any other single word to express their conceptions of Earth, although they could have used "Earth" in the process of setting out their conceptions. To express their conceptions of Earth, they would have had to produce scientific treatises or lectures. The Latin word for Earth is ubiquitous in Copernicus's *De Revolutionibus*, each time expressing the concept of Earth. He expressed that concept hundreds of times in the process of setting out his conception of Earth, something that he did just once (on this occasion). "Earth" expresses part of what is expressed by both "Earth moves" and "Earth stands still." No one's conception of Earth is part of either proposition. On the contrary, the beliefs expressed by those sentences are parts of different conceptions. Our conception of Earth is of course "associated" with our use of the word "Earth." But the mode of association is not expression.

The concept expressed by the word "Earth" does not influence its use in the same way that the associated conception does (Cf. §12.2). Suppose, for example, that contemporaries of Galileo were asked to name a heavenly body supporting life that revolves around the Sun, and were motivated to comply with the request. A Copernican would probably say "Earth," while a Ptolemaist would surely say "There are none." It is the Copernican's conception of Earth which, combined with his desire to comply with the request, motivates him to express the concept of Earth by uttering the word "Earth." The concept of Earth is not a factor *motivating* him to say "Earth." The occurrence of that concept at the appropriate moment was nevertheless a necessary condition of the action. For unless the idea

of Earth had popped into the Copernican's mind between the request and his response, he could not have expressed that idea and consequently would not have used the word "Earth." Of course, it would be difficult to imagine a Copernican not thinking of Earth in response to such a request. But a slight change of the example will illustrate the importance of concept occurrence more dramatically: suppose that we are asked to name a planet with more than one moon that revolves around the Sun. We may very well know how many moons each planet has. But unless the idea of Neptune pops into our mind at the right moment, we will not utter the word "Neptune." Mere possession of an adequate conception of Neptune is not sufficient to produce this response.

In general, the process of recognizing or classifying something as Φ depends on two independent factors: possession of the concept of Φ and its occurrence at the right moment, plus possession of a conception of Φ that is adequate for the subject to acquire the belief that the object is Φ on the basis of the evidence given. Overt behavior, such as saying that the object is Φ, depends in addition on the subject's desires, including the desire to be honest, open, and cooperative.

We distinguished *possessing* (or *having*) a concept from *grasping* (or *understanding*) it in §16.7. To possess the concept of Φ is to be able to reconceive the concept. The thought-part expressed by "Φ" must be capable of occurring to us. To grasp the concept of Φ is to have a certain level of knowledge about Φ. The broader and deeper our knowledge of Φ, the better our grasp of the concept of Φ. To grasp a concept is not simply to have a thought-part, but to have a centered belief system, or conception. Since my conception of quarks is neither broad nor deep, my grasp of the concept is weak. Since my conception of the Earth is extensive and accurate, I fully grasp the concept of the Earth. We cannot speak of degrees of possession. We either have a concept or we do not.

The principle that an idea (concept) cannot occur to us unless all of its components do (Postulate 14.2) does not apply to conceptions. Being systems of belief, conceptions are not event-types. Conceptions are possessed, but do not occur to a subject. We could define "occurrent conception" by analogy with "occurrent belief" (see §12.5), but then we would find that our conceptions of things are invariably too complex to be occurrent. Important parts of the conception may be occurrent, but never anything close to the whole conception. When thinking about astronomy, for example, it may occur to me that Earth moves around the Sun. When thinking about geology, it may occur to me that the surface of Earth is subject to continental drift. When thinking about

language, it may occur to me that Earth is properly referred to by the impersonal pronoun "it." On very few occasions will all three parts of my conception of Earth be simultaneously occurrent. Using the terminology of psychologists, we may say that a conception is *activated* when some, but not all, of the beliefs in the conception are occurrent. In accordance with the laws of association, activation will tend to spread from one component of a conception to another. We observed in §18.3 that the principle that the components of a concept must occur whenever the concept occurs is a consequence of general mereological principles, not of the laws of association.

The principle that we do not have a concept unless we have all of its components does have a close analogue applying to conceptions. For we do not have a particular conception of something unless we have the beliefs constituting the conception. But the similarity here is partly verbal. For "has" has different meanings in "S has the belief that p" and "S has the concept of Φ" (§16.1).

Concepts and conceptions have different kinds of *structures*. The structure of something is defined by the relationships among its parts. Being thoughts or thought-parts, concepts are event-types. Hence their components must occur together whenever the concept occurs. Being systems of beliefs, conceptions are not events. Their components must be jointly possessed by anyone who has the conception, but the beliefs in a system need not be occurrent together, and in large and complex systems they could not be occurrent together. As we observed in Chapter 14, concepts have a *phrase structure*. The relationships defining the structure of a concept resemble the relationships among words in a phrase, clause, or sentence. The structure of a conception, by contrast, is more like the structure of a *theory*. Theories, of course, are expressed by sets of sentences. The relationships defining the structure of a conception are like the relationships among the sentences expressing a theory, and include entailment, incompatibility, and confirmation. Unlike the sentences in a paragraph or story, the beliefs in a conception do not have a sequential order. Conceptions, we might say, have an *inferential system structure*. Deductive systems, like Euclidean geometry, are the most clearly understood types of inferential system. The elements of conceptions more generally can be related inductively as well as deductively. Conceptions also have an *associational system* or *prototype structure* (§19.5). Parts have a tendency to activate other parts, and these links define a structure. We can even view conceptions as having a *conceptual system structure*, since a set of relationships among the concepts contained in the propositions believed is defined by the conception.

The phrase-structure of a concept is an individuating characteristic. Concepts are defined not only by their components but also by the way in which their components are put together. Since the concepts expressed by the individual words "house" and "boat" are related differently in the compound concepts expressed by "houseboat" and "boathouse," the two compound concepts must be different. Similarly, "Vladimir and Artur played the piece" is ambiguous, because it can express either the thought that they individually played the same solo piece, or the thought that they played a duet. A conception, by contrast, is not individuated by its inferential or associative system structure. A conception is defined entirely by the set of beliefs that make it up. A woman's conception of geometric objects does not change if the only difference from one time to another is the order in which she mentally derives the theorems. She might retain a fully Euclidean conception despite a change in her way of inferring the theorems from the axioms. It is particularly clear that the associational structure of a conception is not individuating, since the degree to which two ideas are associated may change owing entirely to what the subject happens to have been thinking about recently, even though the subject's beliefs have not changed at all (Chapter 18). Whether conceptual system structure is an individuating property of conceptions depends on which relations among concepts define the system. If they are restricted to relations created by the beliefs defining the conception, then conceptual system structure will be individuating. If any of the relations result from the inferential or associational system structure, it will not be.

Let us refer to the predicate concepts of the beliefs forming a conception as the "predicates" of the conception. The predicates of a conception need not be, and usually are not, parts of the corresponding concept. This may be so even if everyone shares the conception. It is part of our shared conception of a cat, for example, that cats are vertebrates with stomachs. Nevertheless, it does not occur to us that cats have stomachs every time we think about cats. Since the idea of a cat can occur to us when the idea of having a stomach does not, the latter cannot be part of the former. It is quite possible, in fact, that the concept of a cat is simple or atomic, even though our conception of a cat is very complex. A well-known hypothetical case due to Hilary Putnam suggests that even the concept of being an animal is not part of the concept of a cat. It is imaginable that scientists could have discovered that despite surface similarities between cats and dogs, cats are really amazingly sophisticated automata planted on Earth by Martians. Where we do have a genuine part-whole relationship, such a thing is not imaginable. The suggestion that scientists might have

discovered that red balls are really not red does not even qualify as far-fetched science fiction. Since the concept of being red is not only a predicate of our conception of red balls, but also part of the concept of a red ball, the idea that a red ball is not red is self-contradictory. Fodor (1975, 1981) has argued that natural-language morphemes like "cat" invariably express simple concepts (cf. §15.2). A large part of his evidence consists in the fact that few, if any, strictly analytic definitions have been discovered, despite considerable effort. My point here is not that Fodor's generalization is true, but only that it is not refuted by the fact that our conceptions of the objects denoted by words without grammatical structure are invariably complex.

Conceptions change as beliefs change. Concepts are fixed abstract objects. Man's conception of Earth changed during the scientific revolution. It remained centered on the concept of Earth during the whole period of change. What changed were the propositions containing the concept of Earth that people believed. The concept of Earth cannot change any more than a number could. Nothing at all would count as the concept of a red ball changing into the concept of a blue ball or a red octopus. Given that "the concept of a red ball" refers to the concept expressed by "red ball," the reference of "the concept of a red ball" would change if the meaning of "red ball" changed. But we cannot conclude that the concept of a red ball would change in this case, any more than we can say that the square root of 4 would change if we altered the meaning of the English phrase "the square root of 4." Change of a conception is possible only if the concept that centers it does not change. Otherwise the subject has simply replaced a conception of one thing with a conception of another thing. If S loses the concept of Φ and acquires the concept of Ψ, neither concept changes.[28] S will have "changed his concepts" (i.e., S will possess a different set of concepts afterward), even though no concept he possessed itself underwent a change. (*Analogy*: The phrase "change one's clothes" usually denotes the act of putting on different clothes, not the act of making alterations in one's clothes.)

While concepts do not change, there are moment-to-moment changes in which concepts occur to us. One moment I may be thinking about my son, the next I may be thinking about Earth, and then I may begin thinking about the death of my father. This means that the concept of

28 Contrast Woodfield 1997: 86; and Peacocke 1992: 3. Contrast also Carey (1991: 460, 484), who defines concepts as constituents of beliefs, but nevertheless claims that concepts change as knowledge of their instances is acquired.

my son occurred to me, then the concept of Earth, then the concept of my father. Such changes will account for moment-to-moment changes in what I am talking about, doing, and feeling. During such a period of cognitive flux, there will typically be no change in any of our conceptions. In the period described, my conception of my son, Earth, and my father did not change a bit. I learned nothing new about Earth, and abandoned none of my opinions about my father or my son. (I did manage to refine my conception of concepts.)

Confusions

The confusion of concepts or ideas in our sense with conceptions (or, equivalently, the ambiguity of "concept" and "idea") has been a hindrance to the understanding of concepts second only to the confusion of concepts and images (§19.2). The classical ideational theorists moved freely from "image" to "thought-parts" to "conception," without awareness of the enormous differences. Condillac, for example, begins by defining ideas as "perceptions considered as images of something," and notions as "ideas the mind has framed to itself" (1746: §1.3.16). He later discusses how it came to be that word order in speech does not reflect the order in which ideas are conceived in the thoughts communicated (1746: §2.1.80–101). Finally, he says this:

> It is very likely, for instance, that the notion of gold in the beginning was only that of a yellow and very heavy body: some time afterwards experience made them add the idea of malleability to it, afterwards ductility, and successively all the qualities of which the most able chemists have framed their idea of this substance. . . . Hence the signification of the names of substances must needs have been very uncertain, and been the occasion of a multitude of alterations. We are naturally inclined to believe that others have the same ideas as we, because they make use of the same language. . . . (Condillac 1746: §2.1.113)

When Condillac says that "experience made them add the idea of malleability" to the notion of gold, he must mean "conception" by "notion." He is talking about adding a new belief to our system of beliefs about gold on the basis of experience. Yet in the last line, "ideas" could only plausibly mean "thought-parts." For the fact that Newton and Einstein both spoke English gives us no reason to think that they had the same conception of space and time.

Vendler, who did more than most to clarify the notion of thought, made a less blatant conflation.

To have an idea of a chiliagon is to know what kind of figure it is; to have an astronomical concept of the sun is to know what the sun is in terms of astronomy. Accordingly, the expression of these ideas will amount to a complex definition. This [Cartesian] conclusion is in perfect agreement with our findings in Chapter VI: an idea is an open proposition, often a very complex open proposition. (Vendler 1972: 171)[29]

The agreement is far from perfect. To know what kind of figure a chiliagon is is to know *that a chiliagon is a thousand-sided polygon*. Hence the expression of this knowledge will amount to a definition. But what is known here, and expressed by the definition, is a *closed* proposition, one that is true. An open proposition, by contrast, such as *x is a thousand-sided polygon*, is not the object of any propositional attitude, nor the subject of truth or falsity. Predicate concepts might well be characterized as open propositions. If so, then an open proposition is a thought-part: what is left of a thought when the subject concepts are removed or left undetermined. Knowledge of what something is, by contrast, is a belief or system of beliefs.

One manifestation of the general neglect of thought in favor of belief is the prevalent tendency to overlook concepts (thought-parts) entirely, which leads to ill-fated attempts to assign to conceptions (belief systems) roles that only concepts can play. Consider the following argument from Wettstein as a representative example.

Kaplan, no less than Frege, explains cognitive differences between expressions as differences in their associated modes of presentation. . . . Reflection on some of the anti-Fregean discussions of proper names, moreover, should convince at least these discussants that no such representationalist approach, no matter how benign its representationalism, has any future.

I have in mind here Kripke's Gell-Mann–Feynman case and Putnam's Elm-Beech story. The original point of these examples was to show that reference

29 See also Bruner et al. 1956: 232, 244; Quillian 1968: 227; Vendler 1972: 81, 166; Jackendoff 1983: Chapters 5–8, which weakens Vendler's "definitional constraint"; Linsky 1983: 132, 136, which in some places identifies senses with partial thought determinants, and in others with "criteria of identification"; Evans 1982: 104–9; Fitch 1990; and Wettstein 1991: 155. Historically, see Reid (1785: 383–403, 472), who speaks of concepts being "made up" of attributes, while defining concepts as acts of conceiving objects; J. Mill (1929: 234), who says both that ideas are images and that complex ideas are conceptions, which take together the simple ideas constituting our knowledge of the object; J. S. Mill (1843: §4.2.5), who says that clear conceptions depend on "attentive observation, or extensive experience, and a memory which receives and retains an exact image of what is observed"; and James (1890: Chapter 12, pp. 468–8), who tries to treat change in our conceptions (knowledge) as conceiving new concepts.

does not depend upon what is in the head, upon the information that the speaker associates with the expressions. Don't these examples indicate, no less forcefully, that cognitive significance does not depend upon the associated information? 'Gell-Mann' and 'Feynman', after all, can play cognitively inequivalent roles for a speaker despite the fact that the speaker's associated "conceptual files" are identical. (Wettstein 1991: 155)[30]

Wettstein is describing what has come to be known as the *problem of ignorance* for the thesis that mental or linguistic reference is determined by a subject's conceptions: subjects may well be ignorant of what differentiates one referent from another. There is a parallel *problem of error*: most of our conceptions contain some erroneous beliefs about the referent. Nothing at all had all the properties that the Ptolemaic conception ascribed to the Earth. So for two different reasons, the referent of a term may not be characterizable as the unique object satisfying the conception associated with an expression.

It does not follow that the reference of a term is not determined by the mental representation expressed by the term. "What is in the head" includes concepts (thought-parts) as well as conceptions (belief systems). Both are representations. Hence "cognitive differences" and "differences in associated modes of presentation" are not exhausted by differences in "the information that the speaker associates with the expressions." Even though my conception of the thirty-fifth president is *identical* to my conception of the president who was assassinated in Dallas, the ideas expressed by "the president who was assassinated in Dallas" and "the thirty-fifth president" are different. Hence I can think the thought that Kennedy was the thirty-fifth president without thinking the thought that Kennedy was the president assassinated in Dallas, and others can discover that the thirty-fifth president was the president assassinated in Dallas, or mistakenly believe that he was not. So even if everyone's conception of the thirty-fifth president and of the president assassinated in Dallas were identical, the "cognitive significance" of the terms "the thirty-fifth president" and "the president who was assassinated in Dallas" would be different. Changing from descriptions to names would not alter this conclusion.

One of the roles mistakenly assigned to conceptions is that of meanings in a compositional semantics.[31] Let us ignore the fact that people's

30 See also Salmon 1990: 218–19. Compare and contrast Fodor 1994: 33–4; Rips 1995; Millikan 1998b: §4; Margolis & Laurence 1999: 34–37, 47, 50.

31 For examples of this mistake, see Quillian 1968; Samet & Flanagan 1989: 200–3; Rosenberg 1993: 518. Contrast Fodor 1981: 297; 1998a: 94–108; Osherson & Smith 1981; Jackendoff

conceptions of things differ, so that we can consider the theory that the meaning of a linguistic unit is "the" associated conception. We have already observed some problems with this theory: individual words and phrases do not "express" conceptions; syncategorematic terms are not associated with conceptions at all; and all people's conception of Φ may be identical to their conception of Ψ even though "Φ" and "Ψ" do not have the same meaning. A new problem is that conceptions cannot be combined in the way that meanings are. The compositionality of meaning refers to the fact that the meaning of a linguistic compound is determined by and combines the meanings of its components. Thus the meaning of "lunar whale" combines the meanings of "lunar" and "whale." It seems quite certain that no one's conception of a lunar whale would contain both their conception of whales (living, sea-dwelling, terrestrial, mammalian beings) and their conception of lunar objects (existing on the moon, and therefore nothing but lifeless rock). Similar problems arise even when we confine ourselves to combinations that have referents. Our conception of mammals includes "living all over the globe." But our conception of North American mammals does not include that trait. Our conceptions of fake gold and nongold do not contain our conception of gold. When we consider that our conception of something includes its typical properties (§19.5), the problem explodes: typical pet fish are neither typical fish nor typical pets.

Sentences are particularly problematic on the conception theory, since their meanings seem to differ in kind from the meanings of individual words and phrases. There does not appear to be any corresponding difference among conceptions. If the conception associated with a sentence is taken to be the belief expressed by the sentence, then the conception expressed by a sentence will often be a proper part of the conceptions expressed by its component words, rather than vice versa. For example, the belief expressed by "Reptiles are animals" is part of our conception of reptiles. Another part of our conception of reptiles is that they are cold-blooded. But that conception is not part of the belief that reptiles are animals. It is hard to imagine what a conception theory of semantics could say about how the meaning of a sentence that everyone believes to be false is determined by the conceptions associated with its component

1983: Chapter 8, esp. 139; Schiffer 1990: 263ff.; Fodor & Lepore 1992: 174–86; Rips 1995: 86–98; Kamp & Partee 1995: 168; Cowie 1999: 142–6; Margolis & Laurence 1999: 37–43, 67–8. Contrast also Goschke and Koppelberg 1991: 139–57, who try to retain the identification of meanings with conceptions by denying the compositionality of meaning; and Horwich 1997: 510, fn. 7, who suggests restricting the identification to simple expressions.

words. What operation would take our conception of reptiles (which contains the belief that they are nonbirds) and our conception of birds (which contains the belief that they are nonreptiles) to produce the conception allegedly expressed by "reptiles are birds"?

Concepts (thought-parts) and conceptions (belief systems) are very different. Both are important for psychological theory. Only concepts, however, can serve as the meaning elements in a semantic theory.

§19.5 PROTOTYPE STRUCTURES

Conceptions are often characterized in terms of the set of properties that the subject believes Φ to have. Being at rest is part of Ptolemy's conception of the Earth, it is said, but not of Copernicus's. We have defined conceptions as belief systems, so properties are not literally parts of conceptions. But this property-inclusion terminology is convenient and easily understood: being Ψ is said to be part of S's conception of Φ if the belief that Φ is Ψ is part of S's conception. That is, the following are equivalent.

(11) It is part of S's conception of Φ that Φ is Ψ.
(12) Being Ψ is part of S's conception of Φ.

"Being malleable is part of my conception of gold" is synonymous with "it is part of my conception of gold that gold is malleable." The fact that (12) lacks the repetition of "Φ" in (11) gives (12) the advantage in brevity and style. Characterizing conceptions in terms of sets of properties is natural, given the point made earlier that conceptions are considered the same or different depending entirely on what is believed about the subjects of the conceptions, not on the concepts representing the subjects.

One's conception of something need not consist exclusively of properties believed to be *essential*. Our conception of an object often consists largely of properties believed to be *typical*. Being four-legged is part of our conception of dogs as opposed to spiders, because we know that dogs, unlike spiders, are four-legged. It is part of our conception of a dog even though we know that dogs occasionally lose a leg, for example, or have birth defects. Another part of our conception consists of *instances* rather than *properties*. Thus it is part of our conception of dogs that collies, terriers, and poodles are dogs, along with our pet dog Fido. What is called the *prototype structure* of a conception of Φ consists of all associated concepts of properties and instances of Φ, together with their degrees of association

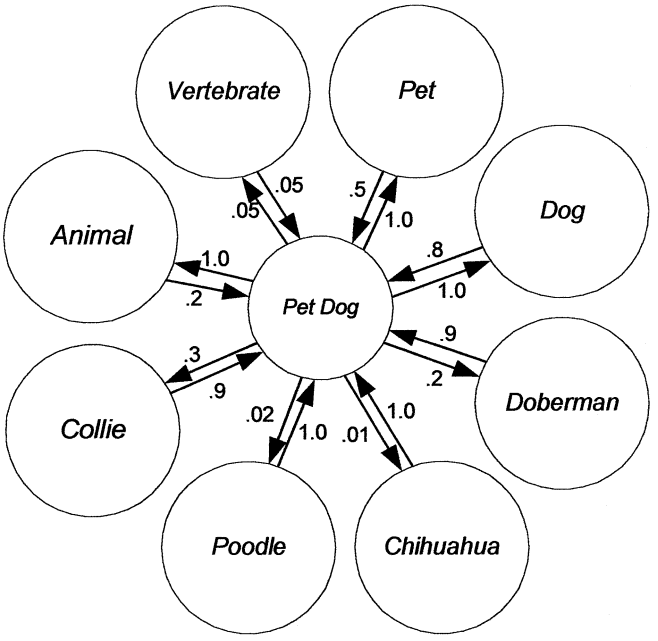


Figure 19.1. Part of the prototype structure of a conception

or other measures of salience or connectedness (see Chapter 18).[32] Figure 19.1 represents part of the prototype structure of a hypothetical subject's conception of a pet dog. The subject whose conception of a pet dog is represented here inevitably thinks "animal" when thinking "pet dog," but seldom thinks "vertebrate." S commonly thinks "collie" when thinking "pet dog," but seldom thinks "poodle." This subject's conception of a pet dog differs markedly from that of a subject who normally thinks "poodle," "dachshund," or "Chihuahua," but seldom "collie," "German

32 See Rosch & Mervis 1975; Smith & Medin 1981; Fodor 1981: 292–3; 1983: 95–6; 1998a: Chapter 5; Jackendoff 1983: §7.5; Stich 1983: 87; Stillings et al. 1987: 65–70, 111–21; Benjafield 1992: 63–4, 70; Kamp & Partee 1995: §6. Bergmann (1982) uses the concept of salience (explicitly) and prototype structure (implicitly) in an interesting account of metaphor, and Lehrer (1974: 10–12) uses it in defining the structure of a semantic field. Related but formally different facets of the structure of a conception are represented by the notion of *frames* in Minsky 1975 and Goffman 1974; *scripts* in Schank & Abelson 1975; *"preference rules"* in Jackendoff 1983: Chapter 8; *"stereotypes"* in Katz 1977b: 39–52, 70–2; *semantic networks* in Collins & Loftus 1975; *full-word concepts* in Quillian 1968 and Johnson-Laird, Herrmann, & Chaffin 1984; and *activity vectors* and *constituent subpatterns* in Smolensky 1988: 66–7. Precursors of the prototype structure idea can be found in Bruner et al. 1956: e.g., 64; Putnam 1970b; and even in Reid 1785: 446–54 (*"regular trains of thought"*).

shepherd," or "Doberman," when thinking "pet dog." A conception is defined as a system of beliefs. A belief system is principally characterized by the set of propositions believed. But conceptions may also differ in the pattern of occurrence of the beliefs, which can be represented as an associative network. Note that because the concept "pet dog" is literally complex, the component concepts "pet" and "dog" will appear in the prototype structure map with maximal association strength.

While the numbers in a prototype structure map represent association strength, they provide reliable predictions of performance on a variety of experimental tasks. For example, when asked to say "True or false?" S is likely to respond to "A pet dog is an animal" faster than to "A pet dog is a vertebrate," and to "A collie is a pet dog" faster than to "A schnauzer is a pet dog." Responses are less likely to be in error when categorizing typical examples. Nevertheless, prototype structure maps are limited representations. They do not indicate how certain each part of a conception is, for example. The subject represented in Figure 19.1 might be just as certain that pet dogs are vertebrates as he is that pet dogs are animals. Moreover, prototype structures do not represent what S's exact beliefs are. We cannot tell from Figure 19.1, for example, whether or not S believes that all collies are pet dogs. Conceptions are complex objects, too complex to be completely represented on any given occasion. In order for some features to be in focus, others must be out of focus.

The associative network map in Figure 19.1 represents part of the structure of *S's conception of a pet dog*, not of *the concept of a pet dog*. From the fact that prototype structures are complex, it does not follow that concepts are complex. This should be patent, since the nodes in associative structure maps represent concepts, while the whole map represents a conception. But the ambiguity of "concept" leads writers like Fodor (1981: 291–3) to speak of the "prototype structure of a concept" (meaning *centered belief system*) in the same breath in which they speak of "the internal structure of a concept" (meaning *thought-part*).[33]

33 See also Samet (1986: 588–92), who says that there is a "tension" in "the folk psychology of concepts" between the idea that concepts are term-sized components of thoughts and the idea that concepts are paragraph-sized bodies of knowledge. Murphy & Medin (1985: 453–4) grapple with the same problem. Margolis & Laurence (1999: 5, 18) similarly distinguish two different models of conceptual structure for "concepts," although it should be evident that the two sorts of structures are possessed by different things. They are led to suggest as possible the theory that the structure of a concept consists in its relations to separate and distinct concepts if they are all organized into a "mental theory" (1999: 47; see also 72–3). The structure of an entity is determined by the relationships among its internal parts, not by its relationships to external objects.

The ambiguity of "concept" has led to looseness in the use of "combination," "composition," "structure," and related terms. Thus E. E. Smith (1988: 35–41) begins with the fact that some concepts such as *red fruit* are composite, and then infers unexceptionably that we must have a way of composing *red fruit* from *red* and *fruit*. To find out how we do this, he conducted studies of prototypes, determining how the prototype structure of *red fruit* is related to those of *red* and *fruit*. Even though it is clear from his results that the prototype of a red fruit is not literally a combination of the prototypes of red and fruit, he describes his experiments as showing "the way in which people combine the prototypes of adjective and noun concepts into the prototypes of conjunctions" (1988: 35). [34]

The same ambiguity leads some to argue that "concepts" might be complex even though they are indefinable.[35] Thus Samet and Flanagan conclude that an indefinable concept may have an internal structure if it has a "prototype representation."

> So, 'diamond' might be represented as:
>
> (13) x is a diamond iff x is of the same kind as the things around here that are used for engagement rings, that are exceptionally hard, that shine, that are expensive, that are weighed in carats, etc.
>
> This is a complex representation, but its elements cannot be construed as parts of the definition of 'diamond.' (Samet and Flanagan 1989: 201)

Now (13) is obviously a complex representation, which is surely part of our common *conception* of a diamond. Nevertheless, (13) is not the *concept* of a diamond, but rather a biconditional proposition *containing* the concept of a diamond as a proper part. Proposition (13) is not a good (analytic) definition precisely because the definiendum "diamond" is not synonymous with the definiens "[object] of the same kind as things around here . . . " Hence even though proposition (13) has all the internal structure of the concept "[object] of the same kind as . . . ," the concept "diamond" does not.

34 See also Rips 1995: 98–9.

35 See Collins & Loftus 1975: 407–9; Jackendoff 1983: Chapters 4–8; Stillings et al. 1987: 67–70; Lakoff 1987; 1989: 106–8; Bechtel 1988: 49; Samet & Flanagan 1989: 201; Goschke & Koppelberg 1991: 137–57; Fodor & Lepore 1992: 186; Silverberg 1992: 124–5; Cowie 1999: 79–80, 147; Margolis & Laurence 1999: 72. A correlative tendency is to use the words "containment" and "structure" quite loosely (cf. §14.5).

Some worry that if "concepts" are literally composed of other concepts and thus definable, then "concepts" should not have a prototype structure.

> The third set of findings consisted of experimental demonstrations that people use nonnecessary features in making semantic categorizations. Taken at face value, these demonstrations constitute strong evidence against the classical view because any model based on the view would presumably be restricted to necessary and sufficient features. (Smith and Medin 1981: 50)
>
> The findings . . . show that not all instances of a concept are equal, yet equality is what we would expect if every instance met the same definition. (E. E. Smith 1988: 22)[36]

The classical view, as Smith and Medin define it, is the view that all instances of a concept have common, defining characteristics. This view holds for all complex predicate concepts. Since the concept *flying animal* is complex, it can be defined: something is a flying animal iff it is an animal that flies. Smith and Medin oppose the classical view to the prototype view, according to which instances vary in the degree to which they share properties. However, the two views are completely compatible. Any defining properties should be parts of the subjects' conceptions, assuming that the subjects can recognize elementary analytic truths. But the instances of a concept will inevitably have, and will be believed to have, properties in addition to their defining properties. Hence our conception of something will almost always contain properties in addition to its defining properties. These nondefining properties need not be common to all instances. Hence our conception of a flying animal can and does have a prototype structure (a bird is more typical than a bat, mosquito, pterodactyl, or flying fish). We should not expect "equality" overall, despite similarity in the defining respects.[37]

§19.6 WHAT IS LEFT

I have now devoted five chapters to clarifying the concept of an idea or concept, enumerating many of the properties of thought-parts and

36 See also Margolis & Laurence 1999: 26–8.

37 E. E. Smith (1988: 26–9) himself points out that even classical "concepts" (e.g., "odd number") have prototype structures, citing relevant experiments. See also Kiel 1989: 29–33 and Margolis & Laurence 1999: 32–3.

distinguishing them from similar yet different entities. Given the history of this subject, however, I fear that the following objection will be going through the reader's mind, despite all that I have said.

> Once you've distinguished ideas or concepts from objects, words, images, and conceptions, I have even less of a grasp of what ideas might be. Once I strip "Aristotle" of the man, the word, the imagery, and all the beliefs I have about Aristotle, there does not seem to be anything left over. Ideas seem otiose.[38]

What is left are all our *occurrent thoughts* about Aristotle, plus their common component, which is expressed by the word "Aristotle." The occurrence of that thought-part constitutes thinking of Aristotle, which is neither a belief nor an image.

In a similar vein, having denied that ideas are composed of sensations or beliefs, it might be asked what ideas *do* consist of. Some ideas, we know, consist of other ideas. But what can we say about the intrinsic nature of the basic or atomic ideas? At present, nothing definite. Introspection enables us to detect various thought-parts, and to distinguish one from another. It enables us to tell that some thought-parts contain other thought-parts. But introspection does not reveal the intrinsic nature of elementary thought-parts, any more than it reveals the intrinsic nature of elementary sensations. Introspection may tell us that our image of an apple contains the sensation of red. But introspection does not tell us what the sensation of red is composed of. Ignorance of their components is no more an argument against the existence of elementary thought-parts than it is against the existence of sensations or subatomic particles. Any form of awareness or method of inquiry has limits, and none could possibly reveal the components of absolutely everything. Given that sensory observation reveals us to be physical beings, and indicates a neurophysiological basis for psychological phenomena in general, there is good reason to assume that ideas (as well as images) are neurophysiological processes. But none of the issues we shall address in this work requires us to take a stand either for or against materialism.

38 This is due to Linda Wetzel and Steve Kuhn, who raised the objection after a brief sketch of my view of ideas. See also Titchener 1909: 182; Ryle 1949: 227; Price 1953: 331, 350–1; Aaron 1967: 201ff.; Schiffer 1978: 197; and Rey 1994: 192. James 1890: Chapter 9, pp. 249–71 and Ryle 1953: 294 strove mightily to reply.

20

The Language of Thought Hypothesis

In §13.1, we said that thoughts are like sentences in being attitudinally neutral, abstract, generatively constructed representations with truth values and sentencelike constituent structures. The analogy between sentences and thoughts has led Fodor (1975) and others to speak of mental representation as a "language," and to speak of the "language of thought."

I believe that the language of thought idea is valuable for two reasons: it takes mental representations seriously, and it stresses their syntactic character. It emphasizes that mental representations have constituent structure. As a metaphor, the notion is unobjectionable, even apt. By providing a useful *model* of mental representation, the language of thought hypothesis serves to demystify the notion of thought, and to suggest ways to study and describe it. But when treated literally, as a serious hypothesis about what mental representations are, the view is misguided. Partly as a consequence, perhaps, there is a considerable lack of clarity in Fodor and others as to how thoughts are supposed to be related to the hypothesized formulas in the language of thought. Are its formulas the thoughts themselves, or "vehicles" that thinkers use in thinking? In this chapter, I will briefly survey the principal forms of the language of thought hypothesis that stretch the sentential analogy too far, and show why they are mistaken.

We will not be concerned with three things that might well be called "language of thought" hypotheses. For one, we will not take issue with the Leibnizian thesis that there is an "alphabet of human thoughts": a system of symbols that perfectly represents the structure of our thoughts, a language with no ambiguous terms or synonyms in which the composition of complex expressions perfectly mirrors the composition of the

ideas expressed.[1] Some claims in this tradition may have been utopian, but there is nothing incoherent in the notion of an ideal artificial language. Indeed, it has been picked up by Frege and developed powerfully by the modern logical tradition.

We can also accept the psycholinguistic thesis that people normally think *in* a language, usually their native tongue.[2] Thinking "in a language" is thinking that is integrated with the production of words that express the thought in that language, either out loud or to oneself. To think in English is to express oneself in English directly and spontaneously, without first using another language and then translating into English. A major step in learning a second language is learning to think in it. The fact that a man is thinking does not entail that there is a language in which he is thinking, so the thesis that we think in a language is substantive. The thesis that we always think in a language is controversial, and probably false (§19.3). But it is common knowledge that we normally do. The fact that we think in a language has few implications for any of the issues with which we are concerned.

Finally, we will not be concerned with the physicalist thesis that thoughts are identical to neural processes, or at least nomologically correlated with them.[3] If physicalism is correct, then the correlation of thoughts with neural processes can be viewed as a *neural code*, in which the neural processes are *natural signs* of the thoughts that they are correlated with. The code in question would resemble the genetic code correlating DNA sequences with genes rather than the Morse code. Neural processes are not expressions and do not have linguistic meaning. The neural network models of association discussed in §18.4 would be one part of the neural code. The neural code thesis is supported by the general success of the physical sciences and the specific evidence currently available from neuropsychology. The code metaphor is apt as long as it is recognized that people do not literally *use* or *know* this code (see §7.8), and do not literally think *in* it. However, the truth or falsity of the neural code view has no bearing on any of the issues with which we are concerned. It neither

1 See, e.g., Leibniz 1709: 7.185; Condillac 1746: 1.216a; and Frege1879. See Kneale & Kneale 1962: 325–30, 479–93, 510–2 and Kretzmann 1967 for overviews.

2 See, e.g., Schiffer 1994: 292. Contrast Bach, who says "it has been argued that people think in their native language . . . I find these arguments . . . far from persuasive"(1987a: 63, n. 11). What Bach really objects to is presented in §20.1. See also §19.3.

3 See, e.g., Schiffer 1982: 144–5; 1991: 188; 1994: 292; Lycan 1990: 277–81. Contrast Dennett 1978: Chapter 2.

implies nor denies, for example, that thoughts have syntactic structure, or that they are representations.

What I *will* criticize in this chapter is the thesis that thought itself *is* a language, and that individual thoughts are sentences of that language. Thought is not literally a natural language, a mental language, a hidden language, or an uninterpreted language. I will be emphasizing ways in which thoughts are *un*like sentences, continuing the demonstration that thinking cannot be treated as a relation to sentences. Positively, I will suggest that thought in the narrow cogitative sense that I have been striving to clarify and distinguish from belief can best serve the theoretical purposes for which a language of thought was postulated.

§20.1 NATURAL LANGUAGE THEORIES

A radical version of the language of thought hypothesis maintains that the thinker's thoughts are sentences of a *natural* language (English, French, etc.). This language is usually that person's native tongue, it is claimed, but may be a second language once the person learns to think in it.[4] On this view, I think in English because my thoughts are English sentences. My thinking is the same act as my uttering words of English out loud or to myself (contrast §12.2).

Harman (1978: 64) offers four points in support of the thesis that the language of thought at least incorporates one's natural language: "(1) that language learners come to be able to 'think in' the new language, (2) that learning a new theoretical terminology seems to involve learning new methods of thinking, (3) that giving arbitrary labels to things helps to structure one's perceptions, and (4) that conversation is often unplanned and involves a certain amount of something it is natural to call 'thinking out loud.'" These points are far from conclusive. There are alternative explanations for points (2) and (3). As for (1), in the sense in which it is common knowledge that we generally think in a natural language, to do so is to formulate or express our thoughts in outer or inner speech, and to do so automatically, without translating. If thoughts were sentences of the natural language we think in, thinking in sentences would entail thinking

4 Harman 1968: 74; 1973: 84–94; 1978; Chastain 1975: 235; Field 1978: 41–2; Devitt 1981: 75–7; Loar 1981: 204–8, 214–15; Devitt & Sterelny 1987: 117–18; Carruthers 1992: 202, fn. 24; Nelson 1992: 183, 243; and Gauker 1994: 25. Contrast Fodor 1975: 56; Bach & Harnish 1979: 138; Stillings et al. 1987: 26; and Fodor & Lepore 1992: 125–6.

in thoughts, and using thoughts to express thoughts. But while we always think thoughts, it makes no sense to say that we think *"in"* thoughts, or use them to express thoughts. Concerning (4), thinking out loud may be unplanned, but even then our verbal utterances cannot be identified with our thinking (§12.2).

One advantage cited for the natural language theory is that it eliminates certain problems concerning the nature of communication. On Locke's theory, public speech must be treated as a "sign" from which private ideas and thoughts are inferred. Since understanding an expression requires knowing what idea it expressed, understanding involves inference. Since thoughts are different from speech, inferences from one to the other are never deductively valid. Hence skepticism is possible as to whether and what our interlocutors are really thinking. This possibility would not exist, it is thought, if the thoughts were the speech itself.[5]

This advantage is an illusion. First, interpretation is neither a deductive process nor a matter of observation. If a speaker says "It ran," we have to know what the speaker was referring to with "it" and what he meant by "ran." We cannot see or hear what the speaker meant, and it is always possible for our interpretation to be in error. Even if the inner-speech utterance "It ran" somehow were a thought, we would not be able to tell which thought without more evidence than the mere fact that the subject was hearing certain words in his inner ear. Second, the inner-speech tokens that are supposed to be the thoughts on this view are distinct from the outer-speech tokens that make up the speaker's public speech. The outer-speech token is still just a sign of the inner-speech token. Inference from one to the other is never deductive. Interpretation is nondeductive even if one holds with Ryle and Sellars that the public speech token is the thought. For there can be no logical guarantee that this public speech token is the thought that it (the cat) locomoted in a certain way (rather than the thought that it [the paint] dripped), or that it is a thought at all (rather than a speech test).

I believe that the natural language version of the language of thought hypothesis is completely untenable, for all of the reasons that led us to distinguish thoughts from sentences (§13.1) and inner speech (§19.3). First, *one sentence may express several different thoughts* ("Flying planes can be dangerous"), and *different sentences may express the same thought* ("Every man is an animal" / "All men are animals"; "Il pleut" / "It is raining"). In

5 Cf. Gauker 1994: 26.

defense of the view that we are criticizing, advocates point to the wide variation found among the tokens of natural language sentences.

> It may be felt that [thoughts] *could not* involve sentences of a natural language: whatever these thoughts are they are *nothing like* such sentences. But what *is* a natural language sentence like? There seems to be no limit to the range of physical types that can yield tokens of such a sentence. In 1.4 we mentioned sound types and inscription types; we might also have mentioned flags, Morse code, sign language, braille, and so on. My stand on the language of thought simply requires that thought types have to be included in that list: tokens in the medium of thought, just like speech tokens and braille tokens, can be tokens of a natural language sentence. . . . (Devitt 1981: 77)

There are limits to the possible tokens of a sentence, however. For example, one thing essential to tokens of the sentence "All men are animals" is that they contain tokens of the words "All," "men," "are," and "animals" in the proper sequence. Consequently, no token of "Every man is an animal" could possibly be a token of "All men are animals." Hence we cannot say that tokens of the thought expressed are tokens of both sentences, and it would be arbitrary to say that they were tokens of one but not the other. *Inner-speech* sentences could well be considered tokens of natural language sentences, despite the radical difference in "medium." But inner speech is not thought (§19.3).

Second, s*ome thoughts are not expressed by any sentence in any natural language of the speaker.* Babies have to think and process information before they can learn any natural language, and many animals think and master cognitive tasks.[6] Moreover, people regularly think thoughts that cannot yet be expressed in any natural language. Finally, suppose that Jack and Jill are both thinking "I went up the hill." Both can use "I went up the hill" to express their thought. But since different thoughts are being expressed, neither can be identified with the sentence "I went up the hill." There do not appear to be any other sentences of English, however, that express their thoughts. "Jack went up the hill" and "Jill went up the hill" certainly will not do, as amnesia shows. This is the well-known "problem of the essential indexical."

Finally, if the thought "p" were the natural language sentence "p," then it could not be true that "p" means "p" because "p" expresses the thought "p." It makes no sense to say that a sentence expresses itself. Opponents of

6 Cf. P. S. Churchland 1980: 148–9, 164; Lycan 1985: 88–9; Maloney 1989: §§4.1–4.2; Weiskrantz 1997.

the expression theory might view this as a virtue of the natural language of thought hypothesis. But at the very least, the fact that on this hypothesis the expression theory is not even an intelligible alternative should give them pause.

§20.2 MENTAL LANGUAGE THEORIES

The counterarguments that we have presented apply only to versions of the language of thought hypothesis that claim that thoughts are sentences in a natural language. These objections can be avoided by postulating that thoughts are sentences in a *mental* language that is different from any natural language.[7] A more general objection looms.

> Representations, mental or not, are symbols, and all symbols need interpretation. Insofar as it is a trail of ink, a sentence is simply a structure devoid of content. It is only as interpreted that a sentence can be said to be contentful. Beliefs certainly have content. So, if they derive their content by way of including mental sentences, these sentences must be interpreted. It is surely the agent in whom such mental sentences occur who must do the interpreting. Interpreting a sentence is itself a thoroughly cognitive enterprise. But how, according to Sententialism, does this process unfold? Must it not itself involve the deployment and interpretation of yet another mental sentence, and this on to infinity? . . . Here there is no final explanation of mentation. (Maloney 1989: 22)[8]

7 See, e.g., Fodor 1975; 1981; 1987; Maloney 1989; Richard 1990: 2; Carston 1994: 4609; Cummins 1997. Contrast Heil 1980; Nelson 1992: 246–8; McGinn 1997: Chapter 6. Some of Boër & Lycan's (1986) informal glosses make it appear that they are offering a natural language theory. Following Sellars, they use "•Tom is a fool•" as a general term denoting any sentence token that "plays the same conceptual or semantic role" that the sentence inside the dots plays for the speaker. When introducing their theory, Boër and Lycan say that they are moved "to follow Ockham and Sellars in regarding thought as a kind of inner speech. . . . we wish to construe John's believing or knowing that Tom is a fool in terms of John's standing in certain relations to •Tom is a fool•'s in his language of thought" (1986: 75). This gloss has all of the difficulties cited in §20.1. But further study makes it clear that on their official theory, believing that Tom is a fool is a relation to the *"logical form"* of the sentence "Tom is a fool," not to the sentence itself. Logical forms come from "a system of 'universal semantic representation'" (56), and are represented by formulas of an artificial language. "A natural language itself is taken to be a set of purely formal rules for projecting logical forms singly or in groups onto the types of marks and noises that are the coin of actual writing and speech. We assume that the semantical properties of a natural-language sentence derive from and are just those of its underlying logical form" (56). It is the logical forms that are the "sentences in the language of thought." So I believe that Boër and Lycan were offering a mental language of thought theory. See also Lycan 1981.

8 Cf. Heil 1980; 1981: 330–42; Lycan 1981: 151; Barwise 1987: 83; Sterelny 1990: 32–3; Laurence & Margolis 1997; Searle 1997: 114–16; Dennett 1997: 222; McGinn 1997: 92–106. Maloney (1989: 25–33) attempts to solve this problem by postulating that the content of

Laurence and Margolis (1997) attempt to defuse this objection by pointing out, correctly, that there are independent reasons for postulating a language of thought. The hypothesis is offered to explain many other things as well, such as the fact that people can respond to stimuli in novel ways. This response does not quash the objection, however. Its main point is that if we postulate that the interpretation of sentences in one language involves the use of sentences in another language, then we have done little to explain what it is to understand a sentence. For we have said nothing about what it is to understand sentences in the mental language, and we get a regress if we postulate the use of sentences in yet another language in order to explain that.

Fortunately, there is a better explanation of all of the phenomena. Instead of invoking a language of thought, whose formulae are the vehicles of thought, it is simpler to invoke thoughts directly. Moreover, we have direct evidence, through introspection, that thoughts exist. Sentences need to be interpreted because they are *expressions*. Sentences have the meanings that they do only because they express certain thoughts. A language is a system that correlates expressions with thoughts (see Chapter 7). To interpret a sentence is to figure out what it means, what thought it expresses. One can recognize a sentence without having the slightest idea what it means, because the meaning of a sentence is not an intrinsic, identifying feature, but an extrinsic, relational feature.

Thoughts do not need to be interpreted because *thoughts are not expressions*. The analogy between language and thought stops here. Thoughts do not express anything, and thought is not a system correlating expressions with thoughts. We cannot speak of thoughts or ideas as being ambiguous, or as being translations, paraphrases, or synonyms of other thoughts or ideas. We *use* expressions, producing and manipulating them intentionally, which requires thinking *about* them. We do not use thoughts, at least not in the same way, and we do not necessarily think about them. We may control our thoughts to a certain extent by trying to solve a particular problem. But we do not have voluntary control over which thoughts occur to us at a given moment. We recognize our thoughts introspectively, and infer the thoughts of others from their behavior. But we do not *interpret* thoughts. The content of a thought is an intrinsic, identifying

mental sentences is determined by their physical properties. But this simply misses the point: whatever determines the meaning of a sentence, we have to interpret it. Lycan resolves the problem by hypothesizing that it is not the person who does the interpreting, but a homuncular subsystem within the person. Since the subsystem must have a language, the process is still regressive.

characteristic, not something contingently correlated with the thought. We cannot identify a thought we are having without knowing what we are thinking. We do not have to discover the content of a thought by observing what the thought is correlated with and inferring. One of the virtues of the computer analogy for cognitive psychology is that it can demystify the idea that thought can control the behavior of a physical entity by providing a model demonstrating that a syntactically structured system of representations can control the behavior of a complex physical object without that object's interpreting or understanding those representations. The internal states of a digital computer, or of a human, do not need to be expressions in order to produce intelligent behavior.

In defense of the language of thought hypothesis, Lycan claims that it "explains why beliefs have the same standard sorts of semantical properties that sentences have. Beliefs have truth-values, bear entailment relations to each other, and so on" (1981: 144, emphasis deleted). Sentences and thoughts are alike in having the *extensional* semantic properties that Lycan mentions: reference, truth conditions, and entailment relations. But sentences differ fundamentally from thoughts in their *intensional* semantic properties – *sense* for sentences, *content* for thoughts. As Frege observed, expressions like "the morning star" and "the evening star" have the same reference but different senses; both are "semantic" properties. Paralleling this observation, the idea of the morning star and the idea of the evening star have the same extension even though they differ in content. The sense (meaning, intension) of a sentence in a genuine language is something that is a determined by the thought it expresses. Since thoughts do not express thoughts, or anything else, they do not have senses. The reference (denotation, extension) of a word is determined by its sense. Since ideas do not have senses, their "reference" is not so determined. The content of an idea is neither its sense nor its reference, although content is like sense in determining reference. If thoughts and ideas were expressions with meaning, then an application of Definition 7.6 would yield an absurdity. It would follow that the idea μ means μ in the language of thought iff the idea μ expresses the idea μ. But nothing can express itself. While it makes perfectly good sense to ask why the word "blue" happens to have the meaning "blue," it makes no sense to ask why the idea "blue" happens to have the content "blue."

Ideas and thoughts do have representational properties, of course, and as a result can be said to have referents and truth-values. The idea of the sky represents the sky, and the thought that the sky is blue represents the sky as being blue. Hence their extensions are the sky and truth,

respectively. For it to be true that people are able to represent their behavior and its consequences, it suffices that they are able to think about their behavior and its consequences. However, the representational character of a thought or idea differs fundamentally from that of words and sentences. The representational character of ideas is an intrinsic and essential characteristic, whereas the representational character of words is anything but. The word "blue" represents the color blue only because of the conventions of English usage, which associate it with the concept of blue. These conventions could change, and the word "blue" could come to represent the color red, or something entirely different. The idea of blue, by contrast, represents the color blue naturally – by its nature and not because of any usage, conventional or otherwise. The idea of blue consequently could not fail to represent the color blue. An idea that represents something else is ipso facto a different idea. Nothing *determines* what an idea represents in the extrinsic way in which usage determines what a word represents.[9]

The traditional sources of the language of thought idea were clear on all of these points (see, e.g., Ockham, *Summa Logicae I*: §1). There is ample evidence, though, that Fodor thought of "formulae in the language of thought" not as a metaphor for thoughts themselves, but literally as expressions.

> I use the term "formulae" without prejudice for whatever the vehicles of internal representation may turn out to be. At this point in the discussion it is left open that they might be images, or semaphore signals, or sentences of Japanese (Fodor 1975: 32). . . . But representation presupposes a medium of representation, and there is no symbolization without symbols. In particular, there is no internal representation without an internal language (55). . . . We are committed to assuming that organisms capable of learning a language must have prior access to some representational system in which such properties can be expressed (79).

Words, sentences, images, and the like are "vehicles" of representation, "media" in which things are *expressed*. Unlike thoughts, words and images are not things whose occurrence *constitutes* thinking.[10] Thoughts are not

9 This is not to deny, of course, that what a complex idea represents is determined by what its components represent. For the components of something determine, in an intrinsic and essential way, what the compound is. See §14.2. This paragraph also assumes that "represents Φ" is interpreted intentionally rather than relationally; see §6.2 and §6.3.

10 Cf. Vendler 1972: 42–4; Sellars 1969: 107; and Barwise & Perry 1983: 12, 270. In support of the idea that "our inner system is our public language," Devitt and Sterelny (1987: 117) observe that "speech often *seems* to be thinking out loud; thought often *seems* to be talking to oneself." I believe this introspective appearance is veridical: people who talk in a certain

"vehicles" of representation, and do not need such vehicles. Thoughts simply *are* representations.

Fodor's fundamental error, I submit, lies in thinking of "formulae in the language of thought" as things that are *both thoughts and expressions*. Fodor is quite explicit about this on occasion.

> Connected with both productivity and systematicity is a further, apparently perfectly universal feature of thoughts and sentences. Their structures are, in the following sense, isomorphic to the structures of the propositions that they express: *If a thought/sentence S expresses the proposition that P, then syntactic constituents of S express the constituents of P.* (Fodor & Lepore 1992: 176)[11]

In this case, the main point easily could be reexpressed without treating thoughts as expressions (replace "express" with "express/represent" or perhaps "express/instantiate"). But elsewhere, the identification is essential.

> Suppose that there are two Lake Eries (two bodies of water so-called). Consider two tokens of the thought "Lake Erie is wet," one of which is, intuitively speaking about the Lake Erie in North America and one of which is about the other one. (Fodor 1981: 238)

It is stipulated that there are two Lake Eries, that is, two bodies of water called "Lake Erie." Given this stipulation, it follows that there are *two thoughts* expressed by the sentence "Lake Erie is wet." Hence we cannot, without failure of the uniqueness presupposition, refer to *the thought* "Lake Erie is wet."[12] If we feel compelled to think of ideas as expressions, we would have to think of them as unambiguous expressions. But it is better not to think of them as expressions at all. There is of course one *inner-speech sentence* "Lake Erie is wet." But that cannot be identified with either of the two thoughts of which it is a vehicle. It almost looks here as if

familiar way are thinking out loud. But "His speaking was his thinking out loud" does not entail "His speaking was his thinking." See §12.2 and §19.3.

11 See also Carruthers 1992: 106; Fodor 1994: 55; Fodor & McLaughlin 1990: 202; A. Clark 1991: 203; C. Macdonald 1995: 3.

12 Fodor gets really tied in knots on these cases. After stipulating, for example, that we are to "[c]onsider two tokens of *the* thought 'Lake Erie is wet,'" he says he has an inclination to take "these to be tokens of *different* thoughts"! Similarly, he says that "[t]wo people would be having different thoughts when each is thinking 'I'm ill' even if *everything* in their heads were the same" (1981: 331). I presume that thoughts are "in the head" if anything is. Hence if everything in two people's heads is the same then their thoughts are the same. In that case, they could not be having different thoughts. Clearly, two people thinking "I'm ill" are thinking different thoughts even if the very same sentence is present to their "mind's ear."

Fodor thinks of thoughts as expressions because he is forgetting his own arguments against the identification of thought and inner speech. Consider finally:

> In fact, since the behavior of a classical machine is sensitive to the syntax of the representations it computes on, even interchanging *synonymous* – semantically equivalent – representations affects the course of computation. (Fodor & Pylyshyn 1988: 102)

Two expressions are synonymous when they have the same meaning. Having the same reference does not suffice for synonymy, as Frege's classic "morning star" and "evening star" show. To be synonymous, two expressions have to express the same idea. Ideas, by contrast, do not express ideas. They have content, and thus can be thought of as having reference. Two ideas may have the same objective reference. But it is impossible for there to be *two* ideas with the same intentional content. The idea of the morning star is not the same as the idea of the evening star. True synonyms, moreover, such as "premise" and "premiss," express the very same idea. What we should say is that since thoughts and ideas are not expressions, the terms "synonymous" and "ambiguous" and their contraries *do not apply*.

We observed in Chapter 17 that concepts cannot be acquired through the same inductive methods by which the meanings of words are learned. We can learn that the word "vixen" means "female fox" by confirming the hypothesis that "vixen" applies to an object iff it is a female fox. But we cannot acquire the concept of a vixen by confirming any hypothesis containing the concept of a female fox, since the hypothesis cannot be entertained before all of the component concepts are acquired. The fact that Fodor took it for granted that concepts are learned in the same way that words are learned is therefore further evidence that he seriously believed that formulas in the language of thought literally are expressions. How seriously he believed this can be gauged by the admitted absurdity of some of the consequences that he was willing to accept.

> Learning a language (including, of course, a first language) involves learning what the predicates of the language mean. [That involves] learning that they fall under certain rules (i.e., truth rules). But one cannot learn that P falls under R unless one has a language in which P and R can be represented. So one cannot learn a language unless one has a language (Fodor 1975: 64). . . . The language of thought is known . . . but not learned. That is, it is innate (65). . . . It follows immediately that not all the languages one knows are languages one has learned and that at

least one of the languages which one knows without learning is as powerful as any language that one can ever learn (82).[13]

How can we possibly credit a two-year-old, just barely learning English, with knowledge of a language more powerful than English or any other language that the child could ever learn? How could the child possibly have concepts as sophisticated as that of truth, reference, meaning, and rule, which he must possess in order to learn what Fodor claims he does learn? Looked at another way, how could a language with words for all of the amazing discoveries of the last hundred years possibly have been wired into human nature since the dawn of man, with a name for every person, place, and thing that will ever be named?

I admit that these conclusions really may seem scandalous. I should be inclined to view them as a reductio ad absurdum of the theory that learning a language is learning the semantic properties of its predicates, except that no serious alternative to that theory has ever been proposed. (Fodor 1975: 82)

But the theory that learning a language is learning the semantic properties of its predicates is not the source of the scandal. The theory that thought is a language is. And that theory is easily dropped, being either an exaggerated metaphor or a category mistake.

Learning that P falls under R requires the ability to *think* the thought that P falls under R. That thought represents P as falling under R. But thinking "P falls under R" does not entail using words, formulas, images, or anything else to represent P as falling under R.[14] Since the ability to think such thoughts does not entail the use of language, learning a language does not require already knowing a language. People do gradually develop the ability to think, of course. But we do not ordinarily describe the process as learning. Whereas child-rearing manuals give rather specific dates for when children learn to sit up, crawl, walk, and talk, they say nothing about when the child learns to think. It is particularly clear that the development of thought is utterly different from the hypothesis testing and confirmation involved in learning that matter is composed of atoms,

13 Cf. Harman 1968; 1975: 286; Field 1978: 45; Devitt & Sterelny 1987: 139–40, 148; Crane 1990: 192–3; Laurence & Margolis 1997; Fodor 1998a: 123–4. See also Vendler 1972: 139–40; Stillings et al. 1995: 26–7; and Cummins 1997: 537. Contrast Heil 1980; P. S. Churchland 1980: 158; Maloney 1989: xxii, 112–13; Nelson 1992: 248.

14 To forestall a possible objection, thinking that s is P does not entail thinking the higher order thought that the concept of s falls under the concept of P. For if it did, we would be off on a vicious regress, since "the concept of s is subsumed by the concept of P" is itself of the form "s is P." Cf. Bennett 1971: §§1.2–3.

or in learning that words have certain meanings. Developing the ability to think is more like learning *how* to do something than learning *that* anything is the case, but even that term does not apply. While Fodor's claim that every normal human being has an innate language seems absurd, as he concedes, the claim that every normal human being has the innate ability to think is pretty banal.

Fodor observed that cognitive psychology largely accepts what behavioristic psychology rejected: the traditional conception of mind, developed over thousands of years before experimental psychology arose, and acquired by every normal adult in the process of growing up.

> What is *un*traditional about the movement, if I have reconstructed it correctly, is the account of propositional attitudes that it proposes: To have a certain propositional attitude is to be in a certain relation to an internal representation. (Fodor 1975: 198)[15]

It does seem undeniable that to have a propositional attitude is to stand in *some* relation to some kind of mental representation. But unless it is added that the internal representations are *expressions*, there is nothing new in this account. The account is positively tautological when the internal representations are taken to be *thoughts*, specifically, propositional thoughts. To believe that the sky is blue, for example, is to stand in a certain relation (namely, the belief relation) to an internal representation (namely, the proposition, thought, or idea that the sky is blue).[16] Fodor seems to reject this idea.

> Suppose that it is propositions that bear truth values. Then part of our theory of propositional attitudes will have to be an *account* of the relation that holds between me and a proposition when the proposition is one of the ones that I believe. So far as I can see, that relation will have to be mediated by a mental representation (viz., a mental *particular* to which causal properties are ascribed). If that sort of treatment is correct, then the mental representation will, in turn, have the semantic property (not of being truth-valuable but) of *expressing* a proposition. . . . If you think this is an artifact of the model, you need some alternative way of construing the relation that holds between a proposition and someone who believes the proposition – a construal which, in particular, does not rely on the notion of

15 See also Fodor 1975: 75–7; 1981: 30; 1987: 17; Harman 1977b; Field 1978: 18–37; Schiffer 1978: 206; 1982: 137; Lycan 1981: 141; Stich 1983: 29, 78–9, 215; Maloney 1989: 5, 47–51; Crane 1990: 191, 193.

16 Fodor's (1998a: 9) recent claim that there can be mental representation without thought seems true when "thought" means "belief." But it is hard to see how someone who is not thinking the thought that the sky is blue could be said to be mentally representing that fact.

mental representation. I know of none such that seems remotely plausible. (Fodor 1981: 317, my emphasis)[17]

First, as Fodor (1998a: 8) would now agree, a theory of propositional attitudes does not *have to* give an "account" of belief. It could take the belief relation as a primitive, using it to define other propositional attitudes, formulating laws relating it to behavior, neurophysiology, and so on. Second, Fodor's account is not terribly informative. As he put it in the passage cited earlier, to believe P is to be related in a certain way to something that expresses P. More recently (1998a: 8–9), he has said that to believe P is to be related to a mental representation meaning P. While we have an intuitive idea of what it is for a natural language sentence to *express* a proposition or to *mean* something, which we have analyzed in Part II, we have not been told what it is for an "internal formula" to express a proposition or to mean something. Moreover, the account of belief provided will not tell us the essential nature of the relation of believing a proposition unless it is part of the essential nature of internal formulas that they express the propositions that they do or mean what they do. But this seems unlikely if the word "express" or "mean" is appropriate. Finally, it is unclear why, or in what sense, believing the proposition that Paris is in France, say, must be "mediated" by a *particular*. Of course, the proposition or thought that Paris is in France is a particular event-*type*, which has numerous tokens, occurring as it does to everyone who thinks the thought that Paris is in France. According to the first law of occurrence (§12.6), that thought must have occurred at some time to anyone who believes that Paris is in France. But S may believe that Paris is in France at times when no thought occurs.

Fodor (1981: 201) conceded the obscurity of the notion that internal formulas express propositions, but he wrote that off as a problem about propositions rather than internal representations. More seriously, Fodor rejected as incomprehensible the idea that propositional attitudes are unmediated relations to propositions.

My main reason for not saying "propositional attitudes are relations to propositions" . . . is that I don't understand it. I don't see how an organism can stand in an (interesting epistemic) relation to a proposition except by standing in a (causal/functional) relation to some token of a formula that expresses the proposition. (Fodor 1981: 201)

17 See also Fodor 1981: 318; 1990a: 168; 1998a: 7–8, 25; Field 1978: 41; Brand 1984: 115; Böer & Lycan 1986: 77–8; Crane 1990: 191.

I have sketched a way of understanding it. Propositions are thoughts, and thoughts are structured event-types. Individuals stand in the belief relation to a proposition/thought in virtue of the causal or functional properties of its tokens. For example, Kathy now believes the proposition that Paris is in France in virtue of certain relatively permanent effects caused by a token of the thought that Paris is in France when it first occurred to her that Paris is in France, or by tokens that occurred subsequently when she remembered that Paris is in France. But *tokens of a thought cannot be said to express the thought*.[18]

I believe that Fodor may have found the claim that propositional attitudes are unmediated relations to propositions inconceivable because he took the term "proposition" to denote the external states of affairs that make thoughts and sentences true (see §20.5). Kathy clearly does not stand in an unmediated relation to the state of affairs consisting of Paris's being in France. The thought that Paris is in France is a mental representation that has that state of affairs as its extension. On my view, believing that Paris is in France is an unmediated relation to that thought, which puts people in a mediated relation to its extension (§13.5).

§20.3 HIDDEN LANGUAGE THEORIES

Fodor has said that "there is no believing-that-P episode without a corresponding tokening-of-a-mental-representation episode" (1998a: 8). Believing is not strictly speaking an episode, of course. People retain their beliefs over long periods of time, even when they are asleep and not thinking anything. If believing a proposition at any time t involves tokening a mental representation at t, then a mental particular representing the proposition that Paris is in France, for example, must exist continuously for the whole time that that proposition is believed. Continuously existing mental representations would seem to be classic examples of occult entities, since we are not aware of their existence. Fodor has recently embraced this consequence with Lepore, claiming that mental representations are *hidden*, and are seldom if ever objects of belief or intention.

The claim is just that the semantic properties of the things we interpret (speech acts and intentional attitudes) are derivative from the semantic properties of

18 If one holds à la Schiffer and Loar that belief does not entail occurrence (see §12.6), one has even more reason to reject Fodor's theory. On such a view, we would have to say that S believes P in virtue of the causal or functional properties of the tokens of propositions/thoughts suitably related to P.

hidden things (hidden, that is, from us but not from god). (Fodor & Lepore 1992: 129)

Mental representations aren't "used" at all; and we typically have no beliefs or intentions with respect to them. (126)[19]

On this view, mental representations – symbols in the language of thought – must not be thoughts, for thoughts are typically not hidden. Thoughts are among the most introspectible mental states, and are as easy to infer from behavior as any other mental state. We often have beliefs about our thoughts, and we are constantly trying to express and communicate them. Moreover, thought is one of the propositional attitudes, whose properties are supposed to be derived from the properties of the hidden entities. On this hidden language theory, mental representations are purely theoretical entities, unobserved but postulated to explain observed behavior and propositional attitudes. Such theories avoid the difficulties inherent in any attempt to treat formulas in the language of thought as both expressions and thoughts. They do this by denying that the formulas are thoughts. Since the postulated formulas are not sentences in English or in any other language that we know of, they must be hidden. We think the thought that the sky is blue, on this view, because we stand in a certain relation to a hidden formula expressing the proposition that the sky is blue.

Hidden language theories have two difficulties. First, as far as I can see, the hypothesis that we stand in a certain unknown relation to a hidden formula meaning that the sky is blue does not in any way *explain* our thinking that the sky is blue. The hypothesis does not increase our understanding of what it is to think at all. No indication has been given as to what the occult computational processes are, what it is for them to mean something, or how they give rise to the familiar process of thinking. So if the hypothesis in question has any explanatory value, it must lie in the light it sheds on other propositional attitudes.

The second point is that the hidden language theory is bound *not* to be the best explanation of propositional attitudes other than thought. For there is another theory with at least the same explanatory power that is better, namely, the thought theory. (1) Other things being equal, it is better to postulate observable entities than unobservable entities. Compared to the hidden language theory, the thought theory has at least the same

19 Cf. Maloney 1989: 19; Fodor & Lepore 1992: 128; Stillings et al. 1995: 26. Dennett (1997: 221) takes the fact that they are not introspectible to be grounds for doubting the existence of mental representations.

explanatory power and postulates introspectible entities. Anything that can be explained by the hypothesis that we stand in an unknown relation to hidden representatives of the proposition that p can be better explained by the hypothesis that we stand in a known relationship (thinking) to the thought (i.e., proposition) that p. (2) The hidden language theory postulates entities that are syntactically similar to both sentences and thoughts. But the representational properties of the hidden formulas are neither intrinsic, as those of thoughts are, nor derived from those of thoughts, as the representational properties of sentences are. So the "semantics" of the hidden entities is unprecedented and quite mysterious. That is, sentences in the hidden language of thought are supposed to represent things, but not in any way that we know anything about. Other things being equal, it would be better to explain other propositional attitudes in terms of thoughts. For thoughts represent things in one of the observed ways.

In two ways, then, the general preference for the known over the occult favors the postulation of thoughts. The best way to avoid difficulties in attempts to treat mental representations as both expressions and thoughts is to abandon the idea that mental representations are expressions and keep the idea that they are thoughts.

§20.4 UNINTERPRETED LANGUAGE THEORIES

Stich (1983) suggests a version of the language of thought hypothesis in which the symbols of the language are thought of as uninterpreted formulas rather than as meaningful expressions. Stich distinguishes the "Syntactic Theory of the Mind" (STM) from the "Representational Theory" (RTM).

> Cognitive theories which cleave to the STM pattern treat mental states as relations to purely syntactic mental sentence tokens, and they detail the interactions among mental states in terms of the formal or syntactic properties of these tokens. (Stich 1983: 9)

> The basic idea of the STM is that the cognitive states whose interaction is (in part) responsible for behavior can be systematically mapped to abstract syntactic objects in such a way that causal interactions among cognitive states, as well as causal links with stimuli and behavioral events, can be described in terms of the syntactic properties and relations of the abstract objects to which the cognitive states are mapped. More briefly, the idea is that causal relations among cognitive states mirror formal relations among syntactic objects. If this is right, then it will be natural to view cognitive state tokens as tokens of abstract syntactic objects. (149)

The representational theory holds that mental states are relations to some sort of representational entity. Stich argues for a theory whose mental sentences are neither identical to representational states nor correlated with them. They have no semantic properties or content. Since the symbol strings in the uninterpreted language of thought hypothesis are not viewed as expressing thoughts, one of the principal objections raised against previous versions is avoided. The theory faces a dilemma, however. On the one hand, if the sentences of this theory are *distinct* from thoughts, then the uninterpreted language of thought theory appears to be a version of the hidden language theory, and can be accused of postulating occult entities. On the other hand, if the uninterpreted sentences are supposed to *be* thoughts, then the theory entails that thoughts have no content, and no truth value, which is absurd.

Stich claims that "psychological theories have *no need* to postulate content or other semantic properties, like truth conditions" (1983: 185), and that content "has played no role in the cognitive science literature" (193). On their face, these are incredible claims.[20] For one thing, the claim that propositional attitudes are relations to meaningless, contentless formulas seems incoherent. What do we believe if we believe such a thing? Furthermore, the difference between believing that apples are good for you and believing that apples are poisonous is a difference in content, which for Stich counts as a semantic property. Stich seems to be claiming, therefore, that such differences are unimportant. Yet we would make very different predictions about a person who believes that apples are poisonous than about one who believes that they are good for you. And when trying to explain why someone ate an apple, it would be very significant to attribute his behavior to the belief that apples are poisonous rather than to the belief that apples are good for you. It is impossible to maintain that beliefs are scientifically important without maintaining that what people believe is scientifically important. Yet to talk about what people believe is to talk about the content of their beliefs. We may not have to postulate *referential* properties (extensions or intensions – see §6.3 and §13.5):[21] whether it is true or false that apples are poisonous plays no role in the explanation. But we do have to postulate the content that, together with objective facts, determines the referential properties.

20 Cf. Pylyshyn 1980: 161; Stillings et al. 1995: 358–9; Fodor 1994: 50–54.

21 This is what Stich (1983: 164–5) seems to be observing when he invokes the "principle of autonomy," according to which "the states and processes that ought to be of concern to the psychologist are those that supervene on the current, internal, physical state of the organism."

Two very different theories can be discerned in the passages just quoted, however, only one of which is absurd. The strong and absurd view is that propositional attitudes *are* relations to uninterpreted formulas. The weaker view is that propositional attitudes can be *represented by* uninterpreted formulas, which are thus a relative of Leibniz's "alphabet of human thought." What the weak STM theory claims is the following.

S1 There is a mapping **M** from mental states to syntactic objects meeting two conditions: (i) Every belief is mapped to a unique formula, every desire is, and so on; and (ii) Laws can be stated using the mapping which collectively specify the functional role of the mental states.

Condition (i), of course, implies that **M** is a *function*.[22] To explain why Mary left the building, the weak STM theorist envisages explanations like the following: Mary has a desire mapped to F → L and a belief mapped to F; for any X and Y, Mary's having a belief mapped to X while having a desire mapped to X → Y results in Mary's having a desire mapped to Y; and Mary's having a desire mapped to L results in leaving the building (Stich 1983: 174–5). If the weak STM theorist ever says "Mary believes F," he takes it to be short for "Mary has a belief mapped to F," not as a claim analogous to "Mary believes 'It is going to rain,'" as it would be interpreted in standard English.

The considerations advanced two paragraphs back about the importance of content imply that the only mappings that will meet Stich's two conditions are mappings that represent the difference between believing that apples are poisonous and believing that apples are good for you. That is, conditions (i) and (ii) will be satisfied only if **M** maps the belief that apples are good for you and the belief that apples are poisonous to different formulas, say $(x)(Ax \rightarrow Gx)$ and $(x)(Ax \rightarrow Px)$, respectively. Nevertheless, the notion of content plays no explicit role in the sort of theory that Stich would use to explain and predict behavior. That theory would consist entirely of the claim that a specified mapping meets the two conditions described here, together with the set of laws mentioned in condition (ii). Nowhere in such a theory would there be any mention of content. Assuming that S1 is true, Stich's claim that behavior can be explained and predicted without referring to content is sustained (cf. Stich 1983: 221–4).

Does this mean that *psychological theory* has no need to refer to content? No. Consider the question, *Is S1 true?* To answer that, we will have to

22 Loar (1981: 142–4) makes a similar claim and refers to the formulas assigned to beliefs as their *sentential indices*.

know that some mapping meets the two stated conditions. In order to determine that any particular mapping meets the two conditions, we will have to be sure, for example, that the belief that apples are good for you and the belief that apples are poisonous are represented by different formulas. We would also have to know that the belief that apples are good for you and the belief that pears are good for you are represented by similar formulas. We would have no confidence in a mapping that paired the belief that apples are good with $(x)(Ax \rightarrow Gx)$, the belief that pears are good with $-Fb \rightarrow (\exists x) - (Mx \vee Bx)$, the belief that if Bob is not French then someone is neither Mandarin nor Brahmin with $(x)(Px \rightarrow Gx)$, and so on. We will have to know, in general, that *any functionally important similarities and differences in content are represented by the mapping M.* If we knew this, we would know something very important about content. And what we knew would constitute a highly general theoretical statement of psychology.

S1 is plausible only because we firmly believe the following: to specify the content of a propositional attitude is to specify what is thought, believed, desired, and so on; the content of a propositional attitude is an intrinsic, individuating property that determines its functional and causal properties; and finally, what people think, believe, desire, and so on, has a syntactic structure. If you imagine that these principles are false, S1 will seem to be groundless speculation. To get a feel for this, imagine that you are trying to predict the behavior of some unknown mechanical device known to have several internal compartments, and then replace "mental states" in S1 with "internal states," "beliefs" with "states of compartment B," "desires" with "states of compartment D," and so on. We would not take the result seriously.

Consider a further query: *Why is S1 true?* To explain this, we would have to explain why the formulas to which M maps all mental states suffice to represent all functionally relevant similarities and differences in content. In particular, we would have to determine whether the syntactic character of the formulas had anything to do with the success of the theory, and if so, why. We would also have to explain why any formula mapped to a thought would also be mapped to a belief, desire, hope, intention, and so on. And we would have to explain why different mental states were mapped to formulae with common constituents. Assuming that S1 is true, these would be among the central questions of psychological theory. The theory that I have propounded provides an immediate answer.

S2 S1 is true because propositional attitudes are relations to propositions, which have syntactic constituent structures. Differences in content imply differences in syntax.

We would similarly explain why a particular mapping satisfied conditions (i) and (ii) by observing that the mapping assigns to each propositional attitude a formula representing the syntactic structure of its propositional object. It is hard to imagine any other way of explaining the truth of S1.[23]

Given S2, we could regard the formulas in the range of M as individual constants that refer to the propositions whose structures they represent. That is, we could view the formulas as *names* of the propositional objects of the mental states in the domain of M. Alternatively, we could view the formulas as sentence constants expressing the propositions whose structures they represent. In either case, we could literally say that "S believes F" means that S has the belief mapped to formula "F." It would no longer be true, however, that the formulas are uninterpreted.

Stich bases other arguments against the representational theory on his analysis of belief. Simplifying, Stich (1983: Chapter 5) holds that to believe that p is to be in a state like that in which we utter "p." Stich argues that the appeal to similarity inherent in judgments of content means that representational theories must inevitably be plagued by vagueness. I have criticized Stich's analysis in §13.2. If his analysis led to unwarranted charges of vagueness, that would provide an independent reason for rejecting it. But the charges are not groundless. It *is* difficult or impossible in some cases to say whether or not a subject holds a particular belief or has a given thought. Stich's case of progressive memory loss and the familiar facts of child development provide classic examples of "borderline" cases. It is not quite clear that these examples show the concept of belief to be vague. For our theoretical knowledge of belief and our evidence about memory loss and cognitive development are still quite limited. So our inability to judge in these cases may be due to ignorance rather than vagueness. But even if we assume that the concept of belief or content is vague, that does not provide any reason to deny the existence of content. Nearly every concept has some degree of vagueness, even in science.

23 Note that I am not asking why beliefs and desires have the causes and effects that they do. As Stich (1983: 176) correctly observes in his reply to Pylyshyn, content-based theories have no answer to this question either. The question is why we are able to map beliefs, desires, and thoughts to formulas in a way that enables us to predict their causes and effects. That can be explained by the fact that the formulas represent the contents of the propositional attitudes.

Pick any biological term that describes the endpoint of a developmental process, and it will be difficult or impossible to apply the term at some point in the developmental process. The term "human" is a well-known example. "Brain" would serve just as well. Semantics is usually compared unfavorably to syntax, but the latter is filled with vagueness. In any spoken-word token, for example, where exactly does one phoneme end and the next begin?[24] Finally, observe that the weak syntactic theory of S1 does not enable Stich to escape the problem of vagueness. If it is vague what belief someone has, it will be just as vague what formula is mapped to it by **M**. That is, it will be vague what function **M** is.

Stich's main reason for rejecting the representational theory is that he believes some mental states have no content at all. The argument again uses Stich's analysis of believing that p, but can survive its rejection.

> It is not the case that every belief-like syntactic state token can be described as a belief that p. For to count as a belief that p, a state must suitably resemble the state which typically underlies the production of some content sentence in our language. And if the correct cognitive theory does indeed posit a category of belief-like states, then almost certainly there are some belief-like state tokens among the very young, the very ill, or among exotic folk which do not much resemble the belief-like states giving rise to any of our own utterances. (Stich 1983: 223)[25]

First, we could concede that *some* mental states do not have content without seriously diminishing the importance of content in psychological theory. As long as what a person thinks, believes, and desires is *ever* critical to what he does, we will need to countenance a "language of thought."

Second, Stich's argument that if the cognitive theorist restricts his theory to mental states with content, then the mental processes of many subjects will be beyond its grasp is unsuccessful. What "unusual subjects" show is that it is difficult or impossible to describe exactly what some subjects believe using what Stich calls "content sentences." We cannot formulate sentences of the form *S believes that p* that are strictly and precisely

24 Stich concludes that "STM theories can do all the explanatory and predictive work of content-based theories, *and they can do it better*. For syntactic theories are not encumbered with the intrinsic vagueness and the built-in expressive limitations that plague Strong RTM theories" (1983: 182). This is an illusion. To determine that clause (i) of S1 is satisfied, for example, we will often have to decide whether two states are or are not the same belief, which on Stich's view depends on whether they have the same "narrow causal profile." The proper resolution of such issues will often be unclear. Suppose Fido chases a vole into a hole at t, and a mole into the hole at t′. Must M map both belief states to the same syntactic formula? (*Hint*: Voles are a lot like moles.) Cf. Maloney 1989: 126–7.

25 See also Stich 1983: 158, 198, 207.

true when interpreted as opaque descriptions. But that concession does not entail that animal and infant beliefs have no content, or that the mental processes of such subjects are beyond the scope of psychological theories that "couch their generalizations in terms of content."[26] It just means that without having the mental states ourselves, we cannot use ideo-reflexive constructions (§7.6).

As an illustration, let us examine animal beliefs, whose subjects are "ideologically unusual." After detailing the familiar respects in which higher animals are similar to, yet different from, humans, Stich (1983: 104–5) argues plausibly that even when we are inclined to say that our dog Fido believes that the squirrel ran up the oak tree, it is questionable whether Fido really believes that it is a *squirrel* that ran up the tree (as opposed to, say, a prairie dog), or that what it ran up is an *oak* tree (as opposed to an elm). We unhesitatingly accept "Fido believes that the squirrel ran up the oak tree" as a *transparent* description of Fido's belief. What is in question is whether it is accurate as an *opaque* description. As noted in §6.2, an opaque description of what someone believes purports to specify the precise proposition that the subject believes. A transparent description merely seeks to locate the proposition that the subject believes within a class of equivalent propositions implied by what he believes. We noted that there may be cases in which our language provides the resources for a transparent belief description, but not for an opaque description. Fido is a plausible example.

Stich's syntactic theory has no trouble handling Fido's belief, for **M** could perfectly well map it to a distinctive formula *Vtu* mapped to none of the beliefs that humans possess. Stich could go on to note that the canine belief mapped to *Vtu* is very similar to, yet different from, the human belief that the squirrel is up the oak tree, which is mapped to *Urs*. A similar strategy is available to the representational theorist, who can proceed as follows: "Fido believes something that is similar to, yet different from, the proposition that the squirrel is up the oak tree; let us use *Vtu* to designate what Fido believes." Fido will then be "within the scope of the generalization" that says that if a subject desires X→Y and comes to believe X, then the subject comes to desire Y. When describing what unusual subjects think, believe, or desire, we are not restricted to the use of English sentences of the form "S believes that p" (or their translations

26 Compare and contrast Fodor (1994: 51), who speaks in terms of "modes of presentation" and "content" where I speak of what is described by opaque and transparent descriptions of content.

into other languages), as Stich believed (1983: 59, 130). When provided with evidence that a subject believes something unusual, we can *name* what the subject believes and describe it even though we cannot think what the subject believes or express it in our own language. This ability is no different from our ability to name and describe a new planet or disease when given indirect evidence of its existence. As observed before, we know of many thoughts we have that are inexpressible in other languages today, and that were inexpressible even in our own language if we go back far enough. So there is nothing extraordinary about the hypothesis that Fido has some thoughts that are inexpressible in English or any other language. Indeed, we have good evidence that this is the case. Following Churchland (1979: 130), Stich (1983: 215) terms the hypothesis that Fido believes a proposition that is inexpressible in our language "untestable" and "incomprehensible." But Stich holds the same thing: he maintains that **M** maps a formula to a belief state of Fido's that is not the same as any of the beliefs that we can express in English.

§20.5 COMPUTATIONAL THEORIES: CONTENT

Fodor distinguishes between the "representational" theory of mind and the "computational." According to the former, "a mental state is (type) individuated by specifying a relation and a representation such that the subject bears the one to the other" (1981: 226). To believe that the sky is blue, I have claimed, is to stand in a certain relation to the thought that the sky is blue. To think the thought that the sky is blue is to stand in a different relation to that thought. So I am working with a representational theory. According to the computational theory, mental processes are "symbolic" and "formal." "They are symbolic because they are defined over representations, and they are formal because they apply to representations in virtue of (roughly) the *syntax* of the representations" (1981: 226). Syntax concerns the structure of representations, rather than their truth, reference (i.e., extension), or meaning.

Fodor argues that the computational theory is stronger than the representational in a way that may suggest that mental representations must be symbols, implying that my views are not compatible with the computational theory.

Consider that we started by assuming that the *content* of representations is a (type) individuating feature of mental states. So far as the *representational* theory of the mind is concerned, it is possibly the *only* thing that distinguishes Peter's thought that Sam is silly from his thought that Sal is depressed. But, now, if the *computational*

theory of the mind is true (and if, as we may assume, content is a semantic notion par excellence) it follows that content alone cannot distinguish thoughts. More exactly, the computational theory of the mind requires that two thoughts can be distinct in content only if they can be identified with relations to formally distinct representations. (Fodor 1981: 227)[27]

I accept the principle that two mental states can be distinct in content only if they can be identified with relations to formally distinct representations. But I reject Fodor's assumption that content alone cannot distinguish thoughts. It is true that *meaning* alone cannot distinguish *expressions*. For meaning is an extrinsic property of expressions. Thus the first two words of "Bill killed the grass" and "Bill killed the lights" are identical even though they have different meanings. *Content* alone can distinguish *ideas* because it is an intrinsic, identifying feature.

Why does the thought that Sam is silly differ from the thought that Sal is silly? Because the former contains the idea of Sam, the latter contains the idea of Sal, and the idea of Sam is different from the idea of Sal. This difference between the thoughts is *both* a syntactic difference *and* a difference in content. The ideas of Sam and Sal, let us assume, are atomic, and are not composed of any more elementary ideas. Still, they are different elementary ideas, so they differ syntactically, just as the dot and dash of Morse code differ syntactically. In general, *the identity and difference of atomic constituents are basic structural relationships.* The intrinsic properties of the elementary units of a system are syntactically irrelevant except insofar as they make the elementary units different in structurally relevant ways. But because intrinsic differences make atomic constituents different, they are always structurally relevant.

Fodor summarizes his argument with a neat syllogism:

There must be mental symbols because, in a nutshell, only symbols have syntax, and our best available theory of mental processes – the *only* available theory of mental processes that isn't *known* to be false – needs the picture of the mind as a syntax-driven machine. (Fodor 1987: 19–20)

If we insist that syntax consists of the formal properties of *expressions*, then Fodor's premise that cognitive psychology requires depicting the mind as a syntax-driven machine is false. It would suffice to picture the mind

27 Cf. Pylyshyn 1980: 113; Kyburg 1980; Schiffer 1982: 116; Stich 1980; 1983: 29, 48, 128, 149–70, 188; Dennett 1982: 26; Churchland & Churchland 1983: 303; Brand 1984: 116; Stillings et al. 1995: 356–8; Fodor 1990: 22–3; 1994: 8–9, 56, 106–7; Fodor & Pylyshyn 1988: 111–13; Maloney 1989: 1–10, 43–4; Crane 1990: 191; Adams & Fuller 1992: §3; Taschek 1995a. Contrast Barwise 1987; Davies 1991; Searle 1997: 118–20.

as a machine driven by *structured events*, events that can be described as complexes of more elementary events combined in specific ways. Many things other than language have formally describable constituent structures, including music, countless mathematical objects, neural processes – and thoughts. If we allow that syntax may encompass the formal properties of representations with a constituent structure other than expressions, as we should, then Fodor's premise that only symbols have syntax is false.[28]

Although syntax is the study of structure or form, not all syntactic properties are structural or formal. Written languages are unusual in this regard, because all syntactic differences therein are structural differences. Being defined by shape, even letters differ in structure or form. But in spoken languages, this coincidence breaks down. In tonal languages, for example, sounds differing in pitch differ syntactically. But pitch is a qualitative rather than a structural notion. If we developed a code in which differently colored dots were substituted for letters of the alphabet, then the dots would differ syntactically without differing in form or structure. All structural differences are syntactic, but not all syntactic differences are structural. All syntactic differences, however, are structurally *relevant* differences – differences that give rise to structural differences. So the claim that ideas differ syntactically because they differ in content does not commit us to the claim that even elementary ideas have a structure of some sort, or to the claim that different elementary ideas differ structurally. Nor, of course, does it exclude those claims. Avoiding the technical term "syntax," the substantive point is that thoughts and other ideas can differ structurally even though their basic components do not, just as two musical compositions can differ in structure even though simple tones do not.

As we have seen, differing syntactically is not incompatible with differing only in content. When it comes to thoughts and ideas, in fact, differing in content *entails* differing syntactically. *Ideas differing in content must be different ideas* (§15.6). Hence they must be either different elementary ideas, or different combinations of elementary ideas. Fodor thinks that this entailment fails for "mental representations," I submit, because he thinks of mental representations as *things different from ideas that express ideas*. Things different from ideas that express different ideas may have all their intrinsic properties in common. For their content is an extrinsic property. Suppose

28 Cf. Fine's (1989: 235) notion of a "universal abstract syntax." Compare and contrast the generative semanticist's contention that "semantic structures have the same formal nature as syntactic structures, except for having semantic rather than morphological entities as their ultimate constituents" (McCawley 1994: 1398).

that symbol A expresses idea X in system 1, while symbol B expresses a different idea Y in system 2. It does not follow that symbol A differs intrinsically from symbol B. They can be one and the same symbol. This occurs, for example, when A and B are the written word "rot," system 1 is English, and system 2 is German. Ideas, however, cannot differ in content without differing in their intrinsic properties. For the content of an idea is an intrinsic property. An idea, unlike a word, cannot have one content in one system and another content in another system.

As is customary, Fodor takes syntax to be the study of compositional structure and form. He is led to deny that differences in content constitute syntactic differences because he opposes syntax to semantics, and assumes that "content is a semantic notion par excellence." Semantics was traditionally defined as the study of meaning, including reference as well as sense. I have argued that the content of a thought or idea is not in any conventional sense its meaning, truth conditions, or reference. I do not count content as a semantic notion because it differs so significantly from meaning, truth, and reference. Whereas the content of an idea is an essential, intrinsic property, the meaning, truth, or reference of a symbol is an inessential, extrinsic property. I do not wish to quibble about the definition of a technical term like "semantics." But if the similarities between meaning and content lead us to take content to be semantic, then we have to give up some standard ideas about semantics. The idea that "semantic" differences never entail syntactic differences will fail. We have to allow that ideas with different contents ipso facto differ syntactically. We also have to abandon the idea that "semantic" properties are always extrinsic and inessential. The content of an idea is an intrinsic property, which cannot be changed without changing the identity of the idea, and which cannot in any sense be lost.

Fodor's conclusion that thoughts are symbols is motivated not only by a narrow conception of syntax, but also by an overly "coarse-grained" conception of content. Another of Fodor's reasons for thinking that psychological mechanisms must be syntactic is this:

> If thoughts have their causal roles in virtue of their contents per se, then two thoughts with identical contents ought to be identical in their causal roles. And we know this is wrong; we know that causal roles *slice things thinner* than contents do. The thought that – – P, for example, has the same content as the thought that P on any notion of content that I can imagine defending; but the effects of entertaining these thoughts are nevertheless not guaranteed to be the same. (Fodor 1987: 140)

If two thoughts have identical contents, then to think those thoughts would be to think *the same thing.* They would be the same thought rather than different thoughts. *Different ideas must differ in content*. Since the thought that – – p is not the same as the thought that p, what we are thinking when we think those thoughts must be different. Since "what we are thinking" denotes what I am calling the content of the thought (§15.6), the two thoughts differ in content. Taken together, the principles "Ideas differing in content must be different ideas" and "Different ideas must differ in content" add up to the familiar principle that *ideas are "individuated" by their content*.

Fodor denies that beliefs are individuated by their content, I contend, because he defines a "proposition" or "content" relationally as the *objective state of affairs* that would make the belief true (see §13.5).[29] Fodor recognizes that Oedipus may believe that Jocasta is eligible without believing that his mother is eligible, so he postulates different "ways of believing" the same proposition. To believe a "proposition" in a certain way, on Fodor's account, is to stand in a certain relation to a "vehicle" whose extension is that state of affairs. To make this account work, "vehicles" have to be thoughts or propositions in my sense, not symbols. Taken intentionally, believing that Jocasta is eligible differs from believing that his mother is eligible because what is believed is different. The proposition believed differs because the concept of Jocasta is different from the concept of Oedipus's mother, despite having the same objective referent.

In short, the computational theory, according to which the mind is a syntax-driven machine, does not require treating mental representations

29 See Fodor 1990a: 167–72; 1994: 39–49; 1998a: 12–15. It is very difficult to deviate from a well-entrenched standard usage for long without lapsing into inconsistency. For example, Fodor later claims that concepts have their contents essentially (1998a: 120). But different objective objects could cause their tokenings in the characteristic way, or none at all. "Symbol-world relations" are not essential. Furthermore, Fodor's "First Thesis" states that "[t]he laws that psychological explanations invoke typically express causal relations among mental states that are specified under intentional descriptions; viz., among mental states that are picked out by reference to their *contents*" (1998a: 7; emphasis changed). In line with the First Thesis, Fodor would grant that the belief that Mark Twain drank water has different causes and effects than the belief that Samuel Clemens drank H_2O. Yet the two beliefs have the same content in Fodor's relational sense. And the belief that Santa Claus brings presents has causes and effects even though it has no content in Fodor's sense. Fodor really gets into a muddle when he later identifies modes of presentation with mental representations (1998a: 21–2), despite having introduced modes of presentation as what distinguishes different but coreferential mental representations. The more standard usage of Fodor's First Thesis is the one I have adopted.

as expressions or as relations to expressions. It requires only that representations are structured, and individuated by their content, in the sense in which content is an intrinsic, essential property. Hence a theory in which mental representations are thoughts rather than expressions can be fully computational. Ideas are individuated by their content, and identity and difference are basic syntactic relationships.

§20.6 COMPUTATIONAL THEORIES: PROCESS

Fodor also appears to believe that the computational theory is committed to an overly narrow view of the cognitive processes that mental representations are involved in. Fodor advocates the Chomskyan "rules and representations" approach to cognitive science, according to which cognition consists in the manipulation of symbols in accordance with formal rules. Fodor appears to hold that the rules must be *rules of warranted inference*. Fodor believes that there is an assignment of formulas to states of an organism on which the "causal sequences that determine propositional attitudes turn out to be *derivations*."[30] He believes that this explains how it is possible for mental processes to be reliably *truth-preserving*. On Fodor's view, "thinking is computation," and "computation is just causation that preserves semantic values."

> Roughly, if you start out with a true thought, and you proceed to do some thinking, it is very often the case that the thoughts that the thinking leads you to will also be true. This is, in my view, the most important fact we know about minds; no doubt it's why God bothered to give us any. (Fodor 1994: 9)

We can agree that inference is a vitally important mental process, and that the human species would not long have survived if human reasoning had not been truth-preserving enough of the time. Since inference is a cognitive process distinct from associative activation, it follows that association psychology's infamous attempt to reduce all cognitive processes to association was doomed to failure.

Fodor's rational psychology errs just as badly. Not all inference is warranted. People often draw conclusions unjustified by their premises. Every reader of this book (especially logic teachers) can surely recall instances of fallacious reasoning. This anecdotal evidence has been confirmed in

30 Fodor 1975: 199, my emphasis. See also Fodor 1975: 73–4; 1987: 18; 1994: 8–9; 1998a: 10–11. Compare and contrast Maloney 1989: xxi, 16–17, 216.

the laboratory.[31] My point here is not simply that the computational rules must allow exceptions, but that there must be rules other than the rules of warranted inference that account for unwarranted inference.

More importantly, not all cognitive processing is inference.[32] Associative activation, recollection, daydreaming, and the imagination that goes on in novel writing and reading do not even resemble inferential processes. The author or reader of a novel in which Hitler rises again might well do some thinking in which the true thoughts "Adolph Hitler was born on April 20, 1889; history records that he committed suicide on April 30, 1945" are followed by the false thought "But he actually escaped to Argentina." Such a thought process is as commonplace as it is unproblematic. Furthermore, since inference is a causal relation among mental states, the basic psychophysical processes such as concept formation could not possibly be inferential processes. For basic psychophysical processes, by definition, relate mental processes to nonmental processes (see §17.3). Finally, many inferencelike cognitive processes are not even supposed to be truth-preserving. Consider the "practical reasoning" by which a woman who wants to increase her earning power and believes that she will increase her earning power if she acquires an MBA forms the desire to acquire an MBA. Formally, B(M → E) plus D(E) causes D(M). Far from being a valid derivation, the argument that her practical reasoning embodies (M → E, E, ∴ M) is an instance of the fallacy of affirming the consequence. Practical reasoning is a syntactically driven process. But the causal sequences involved are not derivations.

Fodor's version of the "rules and representations" view is unusually restrictive in its emphasis on derivation and truth preservation. The more common interpretation specifies only that the *dynamical* rules governing cognitive processes are *purely formal*.[33] Cognitive processes are thus assumed to be like the rules of inference of formal logic, the transformational rules of Chomskyan grammars, or the programs of digital computers. On this interpretation, cognitive processes are *exclusively* syntax-driven: the *only* causally relevant feature of a representational mental state is the syntactic structure of the representation. I shall call this the *formalist theory* of the mind.

31 See, e.g., Woodward & Sells 1935; Tversky & Kahneman 1973; 1974; Johnson-Laird & Steedman 1978; Nisbett & Ross 1980; Stillings et al. 1987: 125–7; Benjafield 1992: Chapter 9; J. Evans et al. 1993. For a careful critical analysis, see Goldman 1986: Chapter 14.

32 Cf. Sterelny 1990: 30. Contrast Maloney 1989: xxi, 51.

33 See, e.g., Tienson 1991; Bechtel 1988: 35.

The formalist theory is stronger than the representational or computational theory, and is, I submit, clearly false if the representations that we are concerned with are ideas and thoughts, so that the set of representational mental states includes the propositional attitudes.[34] It is obvious that basic psychophysical processes cannot be purely formal, for one end of such processes is not a representational mental state. But the formal theory fails even when we confine our attention to processes among propositional attitudes. Let $\Phi(P)$, $\Psi(P)$, and $\chi(P)$ be propositional attitudes, and let P be a propositional mental representation. Whether or not propositional attitudes $\Phi(A)$ and $\Psi(B)$ cause $\chi(C)$ depends on much more than the formal relations among A, B, and C. First, the identity of the particular attitudes Φ, Ψ, and χ is critical. Thus the *belief* that I will lose money if I play the slot machines plus the *belief* that I will play the slot machines would typically cause me to *believe* that I will lose money. But the same conditional belief combined with the *desire* to play the slot machines would not typically cause me to *desire* to lose money. Second, other factors may also be relevant. Imagine three men who believe that all Asian Americans are gifted at math and that Kim is poor at math before they learn that Kim is an Asian American. The first may revise his opinion that all Asian Americans are gifted at math; the second may revise his opinion that Kim is poor at math; and the third may not know how to resolve the contradiction. The classical laws of association also suffice to refute the formalist theory, since recency and frequency of co-occurrence are nonsyntactic factors influencing whether one idea activates another. Even in the simple case of belief causing belief, the times at which the beliefs are possessed is a highly relevant but nonsyntactic factor.

Fodor took formalism to extremes in another way with his view that "having a propositional attitude is being in some *computational* relation to an internal representation" (1975: 198). "Computational processes *result in* S's believing something" is at least plausible. "S's believing something *consists in* a computational relation" commits a category mistake. Assuming that computations are defined as anything like "causal relations among symbols which reliably respect semantic properties of the relata" (1998a: 10), it follows that a *person* cannot stand in a computational relation to anything. Fodor's dictum that "[t]hinking is computation" (1998a: 9) may hold for thought processes, such as those involved in reasoning. But the dictum does not seem applicable to the simple act of thinking a thought. The occurrence of a thought to a person is not a formal relation between

34 Cf. Barwise 1987.

the person and the thought. The same goes for belief and all of the other propositional attitudes.

I have tried in this chapter to dissociate the constituency thesis that I have relied on extensively and argued for carefully from tendentious nominalist, rationalist, formalist, and intellectualist theories that have become associated with the constituency thesis in the language of thought literature. Thoughts and ideas are wordlike, but are not words. The content of an idea is an intrinsic, identifying property, not a contingent, derivative property like sense or reference. Mental processes are syntax-driven, but are not purely formal. Propositional attitudes are relations to mental representations. But the representations are not expressions but thoughts. Thoughts are not hidden, occult entities, but familiar and highly introspectible mental events readily detected on the basis of behavior.

Part Four

Ideational Theories of Meaning

21

Objections to Ideational Theories

I have elaborated the view – for both speaker meaning and word meaning, and for the vast bulk of expressions – that *meaning consists in the expression of ideas*, and that *meaning is determined by the idea expressed*. The word "dog" means something, for example, because it is conventionally used to express an idea. And it means "dog" rather than "cat" because it expresses the idea of a dog rather than the idea of a cat. A particular speaker means "dog" by a word when he or she uses it to express the idea of a dog. Hence the view that I have been developing is an *ideational* theory of meaning. I have acknowledged that for a relatively small set of expressions – interjections and the like – meaning consists in the expression of mental states other than ideas (see §2.5, §7.3). So, more generally, I have been developing a mentalistic theory of meaning. I have concentrated on words for which the relevant mental states are ideas, and shall continue to do so.

A major lacuna in classical ideational theories was that after the identification of meaning with the expression of ideas, little was said about expression. Yet without a detailed understanding of the relation of expression, the tautological character of formulas like "e means μ iff e expresses the idea μ" will wrongly suggest that ideational theories are vacuous. "Saying that the word 'green' expresses the idea 'green' is just a long-winded way of saying that 'green' means 'green,' " the critic could reasonably suggest. The obviousness of the theory will then appear to be a defect.[1] Following the general lead of Grice, Lewis, and Fodor, while striking out in new directions, I have tried to fill the gap. For a speaker to express an idea

1 Cf. Fodor & Katz 1964: 12; Bennett 1971: 3; Dummett 1973: 2; Avramides 1989: 4; and perhaps Quine 1961: 21–2.

is for the speaker to perform an observable act in order to provide an undisguised indication of the idea's occurrence (Chapter 3). For a word to express an idea in a language like English is for it to be conventionally used by English speakers to express that idea (Chapter 8), or for its components and structure to be used by them to express certain ideas and idea structures (Chapter 10). A language is a system of ways of expressing ideas (Chapter 7). English is the system of expression defined by the evolving conventions of a certain community of language users who pass their conventions on from generation to generation (Chapter 11). The expression theory of word meaning that I have presented is both a modification of Gricean use theories and a development of the Aristotelian or Lockean form of ideational theory.

A more serious defect of traditional formulations of the ideational theory was the identification of ideas with either images or conceptions. The images and conceptions associated with an expression are important determinants of its usage. Nevertheless, meaning is not determined by either, and does not consist in their expression. A word like "cat" does not *express* either an image or a conception. And senses do not stand in a one-to-one correspondence with either type of mental representation. We cannot maintain that the images or conceptions associated with complex phrases are composed of the images or conceptions associated with their parts. None of these objections to ideational theories stand when ideas are defined as thought-parts. Part III was devoted to setting out the theory of ideas, clarifying the notion of thought as a propositional attitude distinct from belief and desire, and establishing the existence and nature of thought-parts. I hope I have put to rest permanently the objection that talk of ideas is vacuous, or that it is too obscure or poorly understood to be used in a satisfactory theory of meaning.

Given the tautological character of formulas like "to mean μ is to express the idea μ," it is hard to imagine an alternative to an ideational theory. How could it fail to be true that for "rot" to mean *red* is for it to express the idea *red*? Nevertheless, ideational theories have been severely criticized. The attack resembles the Allied assault on Normandy, with a dozen principal landing groups. The assault has been so successful that ideational theories are seldom taken seriously, unless disguised in putatively innocuous jargon. Some of the objections, I will show, are directed at theses commonly put forward with ideational theories that are not entailed by the thesis that meaning consists in the expression of ideas. These objections show how ideational theories should *not* be formulated. Arguments directed against the core of ideational theories are uniformly

bad. I do not understand why they have been given such credence, but they have, so a response is necessary.

§21.1 THE NON-ENTITY OBJECTION

A defect of what I shall call *Fregean* ideational theories is their claim that *meanings (i.e., senses) are ideas*, and more specifically that

(1) The meaning (sense) of e is the idea μ,[2]

where e means μ. Such a claim is absurd. The idea "water" is a mental event, which occurs to people and has various causes and effects. The meaning of "water" is a linguistic feature of the word and has a very different etiology. The idea "water" is introspectible; the meaning of "water" is not (see §8.5). The word "bank" has many meanings; it surely does not have any ideas. Ideas can be true or false (e.g., the idea that snow is white);

2 Cf. Arnauld 1662: 90; Reid 1785: 394, 477; James 1890: Chapter 12, p. 472; Frege 1892a: 43; 1918: 4–5; 1979: 125; Husserl 1900: 284–6; 327; Hampshire 1939: 5, 19; Russell 1940: 219; Humphrey 1951: 42; Church 1951; Quine 1959; 1961: 21–2; Kneale & Kneale 1962: 497–602; Alston 1964a: 11–12, 20; Katz 1966: 154–5; 177–85; 1972: 37–9; Hamlyn 1967b: 140; Quillian 1968: 223; Hacking 1975: 19, 43–4; Lyons 1977: 254, 316; Devitt 1981: §§5.1, 5.5; D. W. Smith 1981: 102–3, 106; McGinn 1982: 217; Peacocke 1986: 116; Devitt & Sterelny 1987: 56; Jackendoff 1989: 73; Recanati 1990: 705; Böer 1995: 353–9; Fodor & Lepore 1992: 221; Künne 1997; Fodor 1998a: 2; Boghossian 1998a: 199; Horwich 1998a: 4, 20, 98. Loar's (1981: 219–20) identification of meanings with "sentential indices of beliefs" has even greater difficulties, since it entails that meanings are composed of words, and even have meanings of their own. Frege himself, of course, held (1) only for sentences and thoughts (see §§8.4, 13.3, and 21.3, this volume). Contrast Cartwright 1962: 101–2.

With the possible exception of Arnauld and Reid, the classical authors appear never to have asserted or committed themselves to (1). Locke (1690: III.ii) seems to be identifying meanings with ideas when he says that words "stand for," and are "marks of" ideas, which are their "proper and immediate signification." But I would generally translate Locke as meaning simply that words *express* and *indicate* ideas. Cf. Kretzmann 1968 and Martinich 1981: 354. Hobbes (1655: 1.2.5) also seems to have identified meanings with ideas when he says even earlier that "[n]ames are signs not of things but of our cogitations." But Hobbes is concerned with evidential signs, not word meanings. For Hobbes words are "signs" of thoughts because they are "marks" of thoughts – "sensible things taken at pleasure that, by the sense of them, such thoughts may be recalled to our mind as are like those thoughts for which we took them" (1.2.1). Cf. Hungerland & Vick 1981: Chapters 1–3; Hacking 1975: 20, 46–50; and Kretzmann 1967: 376–8. Arnauld usually says only that "words are distinct and articulate sounds used by man as signs of mental activity" (1662: 99; see also 83, 90), and that "speaking is explaining one's thoughts by signs which men have devised for this purpose" (Arnauld & Lancelot 1660: 41; see also 66). But on one occasion he says, "To say that a written or spoken word means such and such is to say only that our minds entertain the meaning, that is, the idea connected with that word whenever we hear or see the word" (1662: 90). Reid (1785: 394, 477) uncharacteristically fails to distinguish the thing conceived from the conception attached to a word, and identifies the meaning of the word with both.

meanings cannot. The idea that dogs bark is a proposition or statement (§13.3), and thus can be asserted or believed; it makes no sense to say such things of meanings. It also makes no sense to say that meanings are associated, or that one activates another. Meanings, on the other hand, can be conventional or stipulated, which cannot be said of ideas; and so on. In general, meanings and ideas have very different properties, and therefore cannot be identical.

The general error that we have been discussing was not patented by ideational theories. Indeed, many referential theorists have claimed that the meaning of a word is its referent, which is even more absurd. The meaning of "Fido" could not possibly be its referent, since only the latter is hairy! Similarly, traditional property theories of meaning erred by identifying meanings with properties of the referents of terms rather than with properties of the terms themselves. The familiar intensionalist view that the meaning of "hot" is the property of being hot has the same sorts of defects as (1). The property of being hot is possessed by burning coals, for example, and is not a linguistic feature. The meaning of "hot" is a linguistic feature, by contrast, and so is not possessed by burning coals. The meaning of "hot" is conventional, the property of being hot is not. Heat burns hands, meanings do not. The theory that meanings are sets of possible worlds, or functions from worlds to referents in those worlds, fares no better. Sets of possible worlds have members, meanings do not. Meanings are possessed by words, sets of possible worlds are not. None of the entities traditionally proposed can be identified with meanings without committing a category mistake. I call this *Cartwright's Problem* (after R. Cartwright 1962).

Remarkably enough, we have been able to explain what it is for an expression to have a meaning without even attempting to answer the classic philosophical question, "What is a meaning?" Definition 7.4 says that for a word to mean μ_i is for it to express i. Hence to have a meaning is to express an idea (or other mental state). This equates having a meaning with expressing an idea, but does not say that meanings are ideas. So the considerations refuting (1) are not even relevant.

The expression theory is therefore as compatible as any other theory with the familiar claim that the question "What is the meaning of a word?" is unanswerable, because meanings are not *things* of any sort.[3] Consider

3 Cf. Ryle 1957: 256, 262–3, 295; Sellars 1958: 224–5; 1963: vii; Austin 1961: §I; Alston 1963; 1964a: 20–2; 1967c: 237; Heath 1967: 178–9; Stampe 1968: §§1–7; Hungerland & Vick 1981: 72; Cresswell 1985: 56; Schiffer 1987a: 265–6; Chierchia & McConnell-Ginet

"impossible." The meaning of this word is *not possible.* But the italicized phrase in the previous sentence is nonreferential: it does not designate or purport to designate any person, place, or thing. We cannot even say that *not possible* is a meaning. For whereas words have, share, and acquire meanings, it makes no sense to say that they have, share, or acquire *not possible.* There is no such thing as *not possible*. Since it is *true* that the meaning of "impossible" is *not possible,* it seems to follow inescapably that the meaning of "impossible" is not a thing of any sort.

A correlative observation is that the verb "means" is not a relational predicate. For the sorts of reasons discussed here and in §6.3, expressions of the form "e means μ" in conventional English cannot be symbolized *Meμ* in quantification theory. The "μ" is a substitutional rather than an individual variable (§6.4). It is a placeholder standing for a range of substituends, but it has no domain of objects. The instances of "μ" are names or singular terms only in a special case. Although "impossible" means "not possible," we cannot treat "not possible" as the name of an entity in the domain of quantification. Even when the instances of "μ" are singular terms, existential instantiation and the substitutivity of identity fail, as the truth of "'The golden mountain' means the golden mountain" and the inequivalence of "2^2 means 2×2" and "2^2 means the square root of 16" illustrate. We can say the following tautologically:

(2) e means what e means.

What this means is that *e means μ, where μ is what e means.* That is, the expression "what e means" is functioning as a substitutional definite description in (2). Since "what e means" is not a referential singular term therein, the meaningfulness of (2) does not support the claim that "means" is a relational predicate.

It is just as easy to argue, though, that meanings are entities. Meanings are things that can be had, shared, given, acquired, or lost by words, and which can be identical or nonidentical and more or less similar. Surely a word cannot have one or more meanings unless at least one meaning exists. We can say that there is no such thing as the meaning of "zlurg," because "zlurg" is meaningless. But we cannot maintain that there are no such things as word meanings without absurdly claiming that all words are meaningless. Finally, we cannot literally hold that the meaning of a sentence is a function of the meanings of its components unless there is an

1990: 15; Nelson 1992: 14; Horwich 1998a: 8–19. Contrast Chomsky 1975: 235–6 and Stillings et al. 1995: Chapter 10.

ordered n-tuple whose first element is the meaning of the sentence and whose other elements are the meanings of its components. A necessary condition for the existence of an n-tuple is the existence of its elements.

We therefore have compelling arguments that meanings are not things, and compelling arguments that meanings exist. I believe that this antinomy is due to a simple but subtle ambiguity. Phrases of the form *the meaning of e* have two senses, one referential, the other nonreferential. On the nonreferential interpretation, "the meaning of e" means *what e means*. Hence:

21.1 **Definition:** *The meaning$_{nr}$ of e is* μ *iff e means* μ.

The meaning of "impossible," for example, is *not possible*. On the referential interpretation, "the meaning of e" means *the meaning that e has* (or equivalently, *the x which is a meaning and is possessed by e*). The meaning of e in this sense is a *feature* or *property* of e, the property that it has of meaning whatever it means.

21.2 **Definition:** *The meaning$_{r}$ of e is the property of meaning* μ *iff e means* μ.

The meaning of "impossible" in this sense is the property of *meaning "not possible."* To say that "impossible" *has the meaning "impossible"* is to say that it has the property of meaning "not possible." There is no sense in which the word has "not possible," which is its meaning in the nonreferential sense.

Given the ambiguity of its putative direct object, (3) is ambiguous.

(3) e means the meaning of e.

When "the meaning of e" is taken nonreferentially, (3) is tautologous, and has the same meaning as (2). "Impossible" means the meaning of "impossible" because it means *not possible*.[4] When "the meaning of e" is taken referentially, (3) makes no sense. The meaning of e in this sense is something that e has, not something that e means. There is no sense in which "impossible" means the property of meaning *not possible*. Thesis (3) would be true when "the meaning of e" is interpreted referentially only if "means" expressed a relation, but it does not. With some forcing, (3) can also be interpreted as an instance of "e means μ"; but then it says falsely that e means "the meaning of e."

4 Thesis (3) has another nonreferential interpretation, on which it means "e means 'the meaning of e.'" On this interpretation it is ideo-reflexive (§7.6) and false.

On the nonreferential interpretation, "What is the meaning of e?" is read with the phrasing "What-is-the-meaning-of: e?" On the referential interpretation, which is rare outside of philosophy and linguistics, the question is read with the phrasing "What is: the-meaning-of-e?" Austin (1961) observed that it makes no sense to ask "What-is-the-meaning-of: a word?" when the answer is supposed to hold for any old word, not just the word "a word." It does not follow as Austin urged, though, that "What is: the-meaning-of-a-word?" is similarly nonsensical when the answer is supposed to hold for all words. The meaning of a word (any word) is a linguistic property, one that differs from its syntax, spelling, or pronunciation.

The referential sense is mandatory when "the meaning of e" follows "has," "acquired," "share," "gave," "lost," or the identity sign. For example, for "vixen" to have the meaning of "female fox" is for it to have that meaning which "female fox" has, not to have that which "female fox" means (which makes no sense). The nonreferential sense is mandatory when the phrase follows "knows," "learned," "forgot," "look up," "explain," or "specify," or when it precedes "is unclear" or "is indicated on page 23," which is the only sort of context that Austin and his followers considered. For example, to know the meaning of "vixen" is to know what "vixen" means, not to know the property of meaning "vixen" (which again makes no sense). Analogously, to know the square root of 4 is to know what the square root of 4 is, not to know the number 2.

As a result of the ambiguity of its subject term, (1) is seriously ambiguous. For different reasons, it is false on both interpretations. On the nonreferential interpretation, (1) is a paraphrase of *e means "the idea* μ*."* This claim is obviously false for nearly every e and μ. For example, the word "water" means *water* in English, not *the idea "water."* Hence when we say "Water is wet," we are talking about water, not ideas; we are certainly not saying that the idea of water is wet.[5] The class of exceptions that proves the rule are expressions of the form "the idea μ." One expression meaning "the idea 'water'" is obviously "the idea 'water.'" However, this is not the interpretation of (1) intended by the Fregean ideational theorists. For on the nonreferential interpretation, (1) does not entail that meanings are ideas, and we could not fill in the blank in "the meaning of e is ___" with "the meaning of e." Indeed, the singular term following "is" would have to be italicized or placed in quotation marks to indicate that it is being used ideo-reflexively (§7.6). On the intended interpretation of (1),

5 Cf. Mill 1879: 1.2.1.

the "is" is the "is" of identity. Hence (1) says that *the meaning e has = the idea* μ, which entails absurdly that *e has the idea* μ. Since this is not a paraphrase of "e means 'the idea μ,'" the above objection is irrelevant. Thesis (1) is false on its identity interpretation too, as we have seen: meanings and ideas have radically different properties.

The ambiguity of "the meaning of e" resembles the familiar *act-object ambiguity* of terms like "desire" and "fear." What e means is an "object" of meaning, in the intentional sense in which the Greek goddess Venus might be an object of desire, and ghosts an object of fear (§6.3). The meaning e has, by contrast, is a meaning property, just as the desire for Venus is a desire state – an "act" of desire. Meaning "blue" is a meaning, just as desiring Venus is a desire. It is just as absurd to treat "means" as expressing a genuine relation in "'round square' means a square that is round" as it would be to treat "fear" relationally in "John fears a ghost." We cannot conclude that there are no meanings, any more than we could conclude that there are no desires or fears. Thus in the nonreferential sense, a meaning is an *intentional object*. In the referential sense, a meaning is the property of meaning such an object.

The "object" of an idea, in the parallel sense, is more generally termed the "content" of the idea (§15.6). While we cannot equate either what e means or the meaning e has with the idea expressed by e, we can equate what e means with the *content* of the idea expressed by e, in one sense of "content." This gives us another seemingly tautological implication of the expression theory for terms that express ideas.

21.3 **Theorem:** *The meaning*$_{nr}$ *of e is the content of the idea expressed by e, provided e expresses an idea.*[6]

We observed in §15.6 that "the content of idea μ" can be taken referentially or nonreferentially as well, although we did not use those terms. Nonreferentially, the phrase means "what the idea represents." Referentially, it denotes an identifying property of the idea, the property of representing what it does. Theorem 21.3 holds when its subject and predicate terms are both taken nonreferentially. It is a consequence of Definitions 7.4 and 15.7, and is a generalization of facts like the following: since "vixen" expresses the idea *female fox*, what "vixen" means is *female fox*, and the content of the idea expressed by "vixen" is *female fox*; since "impossible" expresses the idea *not possible*, what "impossible" means is

6 Compare and contrast Husserl's view as set out in D. W. Smith 1982a: 194–5; Marty's as set out in Liedtke 1990; McGinn 1982: 217; and Carruthers 1992: 10–13.

not possible, and the content of the idea expressed by "impossible" is *not possible*, and so on. Theorem 21.3 would be a category mistake if its terms were taken referentially. In that case, the "is" would expresses identity, and Theorem 21.3 would falsely say that a certain linguistic property of a word (its meaning) is identical to a certain individuating property of an idea (its content). If one term were taken referentially and the other nonreferentially, Theorem 21.3 would be nonsensical.

In the sense in which a meaning is an entity, it is one of a word's *features*. "Feature" is an informal term for *property*, *attribute*, or *characteristic*. So we can at least place meanings in one of the most general ontological categories. Which feature of an expression is its meaning? That is a question that any theory of meaning should be able to answer. Negatively, we can say that the meaning of a word is a feature distinct from its length, spelling, distribution, frequency, origin, and etymology. Positively, a tautological answer that any theory can give is provided by Definition 21.2: the meaning of e is its feature of meaning μ, where e means μ. That is, the feature of e that is its meaning = the feature of meaning μ. To say that "vixen" has, gains, or loses the meaning "female fox" is to say that it has, gains, or loses the feature of meaning "female fox."[7] It follows from Definition 7.4 that the feature of meaning μ_i is the feature of expressing i. This provides an informative answer to the question, "Which feature of an expression is its meaning?"

21.4 **Theorem:** *The meaning$_r$ of e is its feature of expressing a certain idea (or other mental state).*

A word has, acquires, or loses the meaning of "cat" when it has, acquires, or loses the property of expressing the idea of a cat. The meaning of "cat" differs from the meaning of "dog" because expressing the idea of a cat differs from expressing the idea of a dog. Part II was devoted to explaining what it is for a word to express an idea. So Part II implicitly

7 Theorem 21.4 might be rejected on the grounds that while the meaning μ can be called a meaning, it seems decidedly unnatural to call the feature of meaning μ a meaning. However, I think it would be just as unwarranted to distinguish between meanings and meaning-properties on such grounds as it would be to distinguish between colors and color-properties on the grounds that the latter would not ordinarily be called colors. The color red is undeniably a property of objects, namely the property of being red. Nevertheless, while it is natural to classify the color red as a color, we would not ordinarily say that the property of being red is a color. If it is insisted that there is a difference in both cases, then we could replace Theorem 21.4 with something more complicated, such as: *the meaning of e is the feature standing in relation X to the property of meaning μ*, where X is the relation between red and the property of being red, between triangularity and the property of being triangular, and so on.

defined what word meanings are. The fundamental category mistake of traditional "entity" theories of meaning was not their treating meanings as entities, but their taking meanings to be entities *other* than features of words.

§21.2 THE NAMING OBJECTION

The classical ideational theorists said that words "stand for" and "signify" ideas. Given the contemporary preoccupation with reference in the sense that has become standard since Frege and Quine, and given that the classical ideational theorists used "name" for words quite generally, it is natural to interpret Hobbes and Locke as claiming that words *name* or *refer to* ideas.

(4) Words name (or, more generally, refer to) ideas.

This thesis has been correctly described as "one of the classic blunders in semantic theory" (Kretzmann 1967: 177).[8] Mill is famous for his repudiation of it.

> "A name," says Hobbes, "is a word taken at pleasure to serve for a mark which may raise in our mind a thought like to some thought we had before, and which being pronounced to others, may be to them a sign of what thought the speaker had before in his mind. . . ."
>
> Are names more properly said to be the names of things, or of our ideas of things? . . . The eminent thinker, just quoted, seems to countenance the latter opinion. . . . Nevertheless, there seems good reason for adhering to the common usage, and calling (as indeed Hobbes himself does in other places) the word *sun* the name of the sun, and not the name of our idea of the sun. For names are not intended only to make the hearer conceive what we conceived, but also to inform him what we believe. Now, when I use a name for the purpose of expressing a belief, it is a belief concerning the thing itself, not concerning my idea of it. When I say, "the sun is the cause of day," I do not mean that my idea of the sun causes or excites in me the idea of day. . . . I mean that a certain physical fact, which is called the sun's presence . . . causes another physical fact, which is called day. (Mill 1879: Chapter 2)[9]

Thesis (4) is indeed an obvious blunder, for the reason Mill cites. "John" is the name of my son, not of an idea. Hence "John is a boy" is true,

8 See also Warnock 1953: 72–3; Heath 1967: 179; and Bennett 1971: 23.

9 Cf. Frege 1892b: 61; Warnock 1953: 63, 72–3, 76, 82–5; Peters 1956: 134; Boehner 1964: xxx–xxxi; Kretzmann 1967: 394; Heath 1967: 178–9; and Nelson 1992: 14.

while "John is an idea" is false. But the thesis that the meaning of a word is determined by the idea that it is used to express is not committed in any way to the thesis that words name ideas. *Expressing* an idea is not the same as *referring to* or *naming* an idea (compare Chapters 3 and 6, and §7.6).

Surprisingly, I have found only one author before Mill who clearly committed the naming fallacy: James Mill (1829: 137), John Stuart's father and teacher. Ockham (*Summa Logicae I:* §1) was quite explicit in insisting that words signify things rather than our concepts of things. He used the term "subordinate" for the relation between words and concepts, and claimed only that "spoken words are used to signify the very things that are signified by concepts of the mind." Aristotle, Locke, and Hobbes were less clear, but I find it highly uncharitable to attribute (4) even to them.[10] Their works make much more sense if we take them to mean *expressing* by "symbolizing," "standing for," and "marking," and to mean *indicating* by "signifying." On this reading, they held that words are *signs for* things because they are *signs of* ideas (see §2.1), which is not at all implausible. I find nothing in the texts that does not fit this interpretation, and only one line in Locke seems to imply (4) with any directness.[11] Be that as it may, even if Locke, Hobbes, and Aristotle did wholeheartedly and consistently assert (4), they need not have. The thesis is obviously false; it is not required by their central claims about meaning; and it contributes nothing to their theories.[12]

Whereas it is dubious that Locke or Hobbes held (4) at all, Frege explicitly embraced it for indirect discourse: "In indirect discourse words do not have their customary nominata; they here name what customarily would be their sense" (1892b: 59).[13] Following Davidson (1968), Barwise and Perry termed this a fall from "semantic innocence." They raise an interesting theoretical question: how is it possible for us to describe Sarah's mental state by referring to objects that may be miles away from Sarah, as we do when we say "Sarah believes that Clark is well"?

10 Cf. Peters 1967: 120; Kretzmann 1967: 394; 1968; Hungerland & Vicks 1981: 15; and Martinich 1981: 354.

11 "Besides words which are names of ideas in the mind, there are a great many others that are made use of to signify the *connexion* that the mind gives to ideas, or to propositions one with another" (1690: 3.7.1). Cf. Leibniz 1709: §7.2.

12 In the process of correcting his father, J. S. Mill (1869: 43) committed the related fallacy of holding that names are names of sensations, presented under the influence of phenomenalism. It is especially surprising that users of the Cartesian definition of "idea" did not commit the naming fallacy, since it is a consequence of their absurd view.

13 See also Peacocke 1983; Böer 1989: 188; Schiffer 1991: 183–4; 1994: 282, 289.

A Lockean or Fregean will think we have posed the problem wrongly, for on a non-innocent account there *is* no reference to Clark in our report of Sarah's belief. The statement is embedded in the belief report, and is not about Clark at all. On such an account, it is about Sarah's idea of Clark, or a sense of Clark that Sarah's mind can somehow directly grasp. On our view, however, it is Clark himself who is somehow being used to classify Sarah's state of mind. (Barwise & Perry 1983: 225)[14]

I agree that we are using "Clark" to refer to Clark. We are certainly *not* using "Clark" to refer to the *idea* of Clark. If we were, we would be claiming that Sarah believes that the idea is well, which would be a very strange belief for her to hold. Similarly, we are not using "Clark" to refer to its own sense: we are not claiming that Sarah believes a sense to be well or to have any other property.[15] We mean the same thing by "Clark is well" after "Sarah believes that" as we do when we say simply "Clark is well" (cf. Definitions. 7.5, 7.7, 13.2). Since reference is determined by meaning, it follows that we are referring to the same thing in these two contexts. In this respect, indirect discourse differs markedly from quotation. If we say "'Clark' is a proper name," we have used "Clark" to refer, not to Clark, but to the name he bears. Sentences *express* propositions or thoughts, which are a kind of idea. But sentences are not names in any context, and do not *refer to* thoughts or anything else. To refer to a proposition, we need to embed a sentence in an indirect context of the form "the proposition that

14 See also Davidson 1968; Loar 1972; Barwise & Perry 1983: 175, 197, 204, 225; Perry 1980: 218; 1983: 16–17; Stich 1983: 82, 132; Luntley 1984: 269–70; Salmon 1986: 104; Schiffer 1987b: 456; 1992: 507; Crimmins & Perry 1989; Fodor 1990a: 169; Crimmins 1998: 16–17. Contrast Geach 1980: 90; Forbes 1987; 1989: §2; 1990: 535–6; Jacob 1997: §2.1, §6. For an exposition of Frege's view, see Kneale & Kneale 1962: 498 and Salmon 1989c: 439–42. Russell's argument against the theory that judgments consist of ideas is fallacious precisely because it assumes the Fregean view of indirect discourse: "The view seems to be that there is some mental existent which may be called the 'idea' of something outside the mind of the person who has the idea, and that, since judgment is a mental event, its constituents must be constituents of the mind of the person judging. But in this view ideas become a veil between us and outside things – we never really, in knowledge, attain to the things we are supposed to be knowing about, but only to the ideas of those things" (Russell 1910–11: 221–2).

15 Peacocke takes senses to be "modes of presentation," and uses an expression like "[$self_{Mary}$]" to designate Mary's unique self-presentation. Peacocke sometimes says that "any utterance x of 'I'm cold' says that [$self_x$] is cold at [now_t]" (1981, 193–4; see also 1983: 109); cf. Noonan 1980–1: 102; 1984: 222; Perry 1983: 19–26; Sosa 1983: 319; Schiffer 1987b: 462; Forbes 1989: 469; 1990: 252–3; Bealer 1993a: 46; Recanati 1993: 123. Since [$self_x$] is a presentation/sense, it follows that an utterance of "I'm cold" is saying that a presentation/sense is cold! Peacocke (1983: vii) elsewhere acknowledges that this is "loose language," and Schiffer later corrected his mistake as well (1992: 507); see also Perry 1983: 25; Forbes 1990: 548–51; Jacob 1997: 218, fn. 1; Fodor 1998a: 17–18; Crimmins 1998: 11.

p." In "S believes p," "p" has the same function that it has in "S believes the proposition that p," and in the latter it is clearly "the proposition that p" that refers to the proposition, not "p" or "that p."

There is nothing in an ideational theory, however, demanding the account of indirect discourse that Barwise and Perry rightly reject. The ideational theorist can claim, with complete plausibility, that when we use "Sarah believes that Clark is well," we do the following: we express the idea of Clark, and consequently refer to Clark (see §6.1); we refer to Clark in the process of expressing the thought (but not the belief) that Clark is well (see §2.4); we express the proposition that Clark is well in the process of expressing the thought and belief that Sarah stands in the belief relation to that proposition. That is, we refer to the proposition expressed by "Clark is well" in part by using the sentence to express that very proposition. This is what I called *ideo-reflexive reference* in §7.6. We could proceed differently, by first naming or numbering all of the propositions that we wish to talk about, and then claiming, say, that Sarah believes #569. But the device of ideo-reflexive reference enables us to use the very same linguistic resources that we use so efficiently to express an unlimited range of propositions about the extramental world for the additional purpose of ascribing to people those mental states that consist in relations to such propositions.[16] We do not need to regard ordinary belief statements as involving "pretense" or "fiction" of any sort (*pace* Crimmins 1998) in order to maintain that belief statements are both opaque and "innocent."

Frege was led to his famous claims about indirect reference by a simplistic view of semantic compositionality, according to which the referent of a compound expression is *invariably* determined by the referents of its components. There is no reason for anyone to follow him down this misguided path. The truth of "Sarah believes that Clark is well" does not entail that Clark exists, any more than "Sarah believes that God is good" entails the existence of God. Hence the word "Clark" may have no extension even though the belief sentence in which it appears is true. Linguistic intuition or introspection suffices to establish the facts that Barwise and Perry cite as evidence, namely, that we use "Clark" in "Sarah believes that Clark is well" to refer to Clark. Neither suffices to establish that "Clark" has an extension. Furthermore, even when the terms in sentence "p" do have extensions, the truth-value of "S believes that p" is not determined by those extensions. Thus "Sarah believes that President Kennedy was handsome" is not equivalent to "Sarah believes that the thirty-fifth president

16 Cf. Stich 1983: §3.5; Schiffer 1987: 456; Crimmins & Perry 1989: 710.

was handsome," even though "President Kennedy" and "the thirty-fifth president" have the same referent. Barwise and Perry's observation that "President Kennedy" and "the thirty-fifth president" have their conventional referent in these contexts follows from the fact that the terms are being used in their conventional senses, together with the Fregean principle that sense determines reference. It does not follow that two belief sentences containing these terms must be equivalent, unless one adopts the faulty view of semantic compositionality that Frege also held.

Schiffer (1991: 183–4; 1992: 282) reversed Frege's argument and concluded that natural languages do not have a compositional semantics, starting from the premise that the reference of a that-clause is not determined by the referents of its components. Schiffer's argument, like Frege's, is based on the assumption that a language has a compositional semantics only if the reference of a complex expression is determined by the referents of its components (except for idioms). This assumption is problematic, because it defines compositionality in terms of reference (extension) rather than meaning (or sense). Even though Schiffer is correct about the reference of a that-clause, the *meaning* of a that-clause *is* determined by the *meanings* of its components. That is, the idea *expressed* by a that-clause is a function of the ideas expressed by its components. That is the sense in which natural languages are obviously compositional (Chapter 10).

Forbes describes another problem with the Fregean analysis of indirect speech, which he calls the *iteration problem*. He uses "C(E)" to denote an opaque context C governing a sentence E. On Frege's theory, E refers to its ordinary sense in C(E). Forbes then imagines that C(E) is itself embedded in an opaque context C′(C(E)), which is embedded in an opaque context C″(C′(C(E))), and so on.

> The result is that as we increase the depth of the embedding of E by prefixing more and more opacity-inducing contexts, the reference of E, which starts out as its customary sense when E is embedded to depth 1, changes with each additional prefix to a new sense unique to the new depth of embedding. But this is quite implausible; we have no grasp of the ever more abstruse senses that this account postulates, yet we are able to understand iteration of opaque contexts, as in "I wonder if Lois knows that Ralph is spreading the rumor that the police suspect that Superman is a spy," without difficulty. (Forbes 1993: 311)

The fact is that "Superman is a spy" has the same meaning in "I wonder if Lois knows that Ralph is spreading the rumor that the police suspect that Superman is a spy" that it has in "Lois knows that Ralph is spreading the

rumor that the police suspect that Superman is a spy," and in "Ralph is spreading the rumor that the police suspect that Superman is a spy," and so on. This fact is a simple consequence of Definition 13.2. Since "Superman is a spy" is only used once no matter how deeply it is embedded in iterated ascriptions, it only expresses one proposition and has only one meaning.

§21.3 THE SENSATIONIST OBJECTION

Ideational theories of meaning have long been hobbled by the misguided sensationist thesis that our "inner lives" consist of nothing but things like sensations and images, which the term "idea" must therefore denote if it is to denote anything at all.

> The vague conception, common, for instance, to both the British empiricists and Aristotle, whereby a word represents an "idea," and a phrase or sentence accordingly represents a complex of ideas, is simply too crude to serve even as a starting point; it virtually forces us to adopt the conception whereby the meaning of a word is embodied in a mental image. (Dummett 1967: 228)[17]

It is not conventional usage that forces us to use "idea" for "image." *Webster's Encyclopedic Unabridged Dictionary* defines "idea" in terms of "thought" and "conception," listing "mental image" as an *obsolete* sense. Nothing in psychology forces us to identify ideas with images either. Concepts (thoughts and thought-parts) are markedly distinct from images

17 The immediate move from ideas as "thoughts or concepts" to ideas as "images" also occurs in Titchener 1909; 1914; Warren 1921: 297–9; Bloomfield 1933: 142–4; Ryle 1953: 294; and Allan 1986: 88. Humphrey (1951: 128) says that meaning and thought must be distinguished from any images that might accompany it, but still rejectes the popular thesis that speech must have been preceded by an "idea" on the grounds that it need not be accompanied by images (259). Goodman (1952: 67) equates ideas with images, and concepts with what is conceivable (i.e., possible), overlooking thought-parts altogether. And even Aristotle (*De Interpretatione*: Chapters 1–3) and Descartes (*Meditations*: III) speak of both thoughts and images. The locus classicus is Berkeley's Introduction to *The Principles of Knowledge*. See also J. Mill 1829: Chapter 4; James 1890: Chapter 9, pp. 264–5; Peters 1956: 134–5; 1967: 126–7; Kretzmann 1967: 394, in a discussion of Mill; Danto 1975: 16, in the Preface to the *Port Royal Grammar*; Hacking 1975: 23–4; Posner 1980: 171ff., criticizing Wundt and Titchener; Hungerland & Vick 1981: 75, 89, expounding Hobbes's views (e.g., 1655: Chapter 2); Atlas 1989: 12ff., defending Locke's view of "abstract ideas" as sketchy or undetailed images; and especially Hampshire 1939: 2, 8–17, defending Berkeley's theory of signs. Bennett (1971: §1.4) attributes Locke's "double use" of the term "idea" to "his substantive mistake shared with Berkeley and Hume and others in the empiricist tradition of assimilating the sensory far too closely to the intellectual." See Urmson 1967 for a history of the term "idea," and Humphrey 1951: Chapters 2–4 for a survey of the "imageless thought" controversy in introspective psychology.

(Chapter 19), a fact clearly recognized as long ago as Arnauld (1662: Chapter 1) and probably before. And as reviewed in Chapter 19, ideational theories that explain meaning in terms of concepts are superior in numerous respects to imagistic theories.

Consider, for example, Berkeley's (*Principles*: 1.13) famous attack on Locke's doctrine of abstract ideas.[18] If we say that "triangle" has the meaning that it does because it expresses an image of a triangle, then we have to ask whether the image it expresses is of a scalene, isosceles, or equilateral triangle. If we pick one of these, then the meaning that our theory assigns to "triangle" is too specific. If we insist that these three options are too specific to provide the meaning of "triangle," then we are stymied by the fact that there is no "abstract image" of a triangle – none that is neither the image of a scalene triangle, nor the image of an equilateral triangle, and so forth. Images may be vague and indeterminate in various respects, but they are too particularized to serve as meanings for general terms.

While Locke's own words on the subject of abstract ideas are confused, the problem can readily be solved if we focus on thought-parts rather than images. It is not at all contradictory or absurd to say that the idea (concept, thought-part) of a triangle is not the idea of a scalene triangle, nor the idea of an isosceles triangle, and so forth. Indeed, this is obviously true. From the fact that I am thinking the thought that something is a triangle, it does not follow that I am thinking that it is scalene, that it is isosceles, and so forth.[19]

Consider the Fregean critique as formulated by Avramides (1989: 141).

> Frege argues in particular that "an essential distinction between the idea and the sign's sense" is that the latter may be the common property of many while the former is a part or mode of an individual mind. Very roughly, for Frege sense is public, while ideas are private. This observation led Frege to characterize senses

18 See also Hume, *Treatise:* 1.1.7; J. S. Mill 1865: 393; Warnock 1953: Chapter 4; and Bennett 1971: §1.3, §2.6; contrast James 1890: Chapter 12, pp. 468–73; Chapter 18, p. 49.

19 After savaging the concept of an "abstract idea," Berkeley went on to introduce the concept of a *notion,* and effectively advocated a "notional theory of meaning": "We may not, I think, strictly be said to have an *idea* of an active being, or of an action, although we may be said to have a *notion* of them. I have some knowledge or notion of my mind, and its acts about ideas, inasmuch as I know or understand what is meant by those words. What I know, that I have some notion of. I will not say that the terms *idea* and *notion* may not be used convertibly, if the world will have it so. . . . But if, in the modern way, the word *idea* is extended to spirits, and relations, and acts, this is, after all, an affair of verbal concern" (Berkeley, *Principles:* §142, added in the second edition; see also §§27, 89, and 139; Reid 1785: 189). A more charitable interpreter than Berkeley would have asked whether Locke just might have been using "idea" to mean "notion."

as abstract entities, items of what he calls "a third realm." Items of this realm exist eternally and require no bearer for their existence.[20]

But as we have just explained, the idea of a woman *is* an abstract object that is the common property of many. People around the world think about women hundreds of times a day. The idea of a woman is an event-type that occurs to everyone who thinks about women. The eternality of event-types and other universals is controversial, of course, but it is as plausible for ideas as for any other type of type: after every thinking thing has perished, one thing that will be true is that the idea of a woman will no longer occur to any being.

Frege argued as he did because he identified an idea as "an internal image, arising from memories of sense-impressions"(1892a: 59) – a "mental picture" (1884: §60), a "picture that is called up by the imagination" (1979: 131),[21] in the tradition of Berkeley and Hume. Unlike "thought," "concept," and "idea," the general term "image" usually denotes event-tokens rather than types. Unlike "the idea of a woman" or "the thought that Mary is a woman," "the image of a woman" is incomplete, and raises the question "formed by whom and when?" Two people forming an image of a woman will inevitably be forming qualitatively as well as numerically distinct images. Two people conceiving the idea of a woman are necessarily conceiving the same idea. The facts Frege cites refute the identification of ideas with images, as well as the correlation of images with senses, but not the correlation of ideas with senses.

Consider next the Wittgensteinian critique as formulated by Kripke (1982). Kripke asks what makes it true that a speaker means "green" rather than "grue," or "plus" rather than "quus."

20 Cf. James 1890: Chapter 12, p. 481; Bloomfield 1933: 143; Ryle 1949: 295–6; 1953: 294; Gale 1967: 500; Hacking 1975: 49–52; Evans 1982: 104, fn. 24; Brand 1984: 117; Cresswell 1985: 56; Allan 1986: 88, 139; Chierchia & McConnell-Ginet 1990: 58; Schiffer 1990: 260–1; and Nelson 1992: 52–3. See also Bradley's objections to associationism, which is critiqued in Humphrey 1951: 12, 15–19. Contrast also Fodor 1998a: 17–20.

21 That is, Frege here used "idea" (*Vorstellung*) for what he had earlier called a "subjective idea." "An idea in the subjective sense is what is governed by the psychological laws of association; it is of a sensible pictorial character. An idea in the objective sense belongs to logic and is in principle non-sensible, although the word which means an objective idea is often accompanied by a subjective idea which nevertheless is not its meaning. Subjective ideas are often demonstrably different in different men; objective ideas are the same for all. Objective ideas can be divided into objects and concepts. I shall myself, to avoid confusion, use 'idea' only in the subjective sense" (1884: §27, n.1, quoted in Hill 1991: 59). It should be noted that "objective ideas" are also subject to the laws of association (Chapter 18), but cannot be identified with what words mean (§21.1).

It has been supposed that all I need to do to determine my use of the word 'green' is to have an image, a sample, of green that I bring to mind whenever I apply the word in the future. When I use this to justify my application of 'green' to a new object, should not the skeptical problem be obvious to any reader of Goodman? Perhaps by 'green', in the past I meant *grue*, and the color image, which indeed was grue, was meant to direct me to apply the word 'green' to *grue* objects always.

Wittgenstein does not base his considerations on any behavioristic *premise* that dismisses the "inner". . . . Careful consideration of our inner lives, he argues, will show that there is no special inner experience of 'meaning' of the kind supposed by his opponent. (Kripke 1982: 20, 44)[22]

It may be true that an image of a green object is also an image of a grue object (§19.2). Hence our meaning "green" rather than "grue" cannot consist merely in having an image of a green object. Wittgenstein's mistake lay in overlooking other aspects of our inner lives. My thoughts are every bit as real as my images, and no less introspectible. Whether I am thinking of green or grue is something that I know about myself. The idea of green, moreover, is very different from the idea of grue. Hence an occurrence of one cannot also be an occurrence of the other. While the idea of green has occurred to me daily since childhood, the idea of grue did not occur to me until I took a course in the philosophy of science in college. This is so even though in retrospect I realize that I had often seen grue objects, worn grue objects, and formed images of grue objects.

Blackburn repeats the mistake in a slightly different way. Blackburn rejects the suggestion that the difficulties in Locke's theory are due to interpreting "idea" as anything like an image or iconic representation. Instead, he thinks the problem holds for any "dog-legged" theory, according to which the meaning of a word is determined by the mental state it expresses (1984: 43).

What does the introspective candidate find as he considers his own mind, which tells him determinately which hypothesis is true? Perhaps pictures, or formulae, or definitions of terms. But the presence of any such thing *cannot be the fact* which determines which hypothesis is true. . . . No thing can halt the regress of interpretation, for any thing can be taken in different ways, or in no way at all. (Blackburn 1984: 72)[23]

22 See also Blackburn 1984: 45–50, 69–109; Peacocke 1992: §5.4. An object is grue iff either (i) it is green and has been observed before t (pick any time in the future here), or (ii) it is blue and has not been observed before t.

23 Cf. Travis 1997: §7.

Ideas, thoughts, beliefs, and intentions are all things that cannot be "taken in different ways." The idea of green is not the idea of grue. Blackburn goes on to suggest that the problem is confined to theories on which meaning is determined by "the sudden presence to the mind of a display" rather than by something like an intention (1984: 72–3). But then the problem is not with dog-legged theories in general, but only with imagistic theories. The intention to describe something as green differs from the intention to describe it as grue no more introspectibly (and no less) than the ideas of green and grue and the thoughts containing them.

§21.4 THE PRIVACY OBJECTION

As often noted, Locke held that words are "external sensible signs, whereof those invisible ideas, which his thoughts are made up of, might be made known to others" (1690: §2.2.1). Some of Locke's successors claimed not just that one person's ideas were *invisible* to others, but that they were *unknowable* to others. This led to a famous Rylean objection.

> To understand an expression is not to infer an unwitnessable cause. The very fact that an expression is made to be understood by anyone shows that the meaning of the expression is not to be described as being, or belonging to, an event that at most one person could know anything about. (Ryle 1949: 295)[24]

Since we often know what others mean, and what words mean, meaning cannot involve an *unknowable* entity. It does not follow, however, that meaning cannot involve an *invisible* entity. As modern science and common sense both attest, what we know is not limited to what we can see, hear, or otherwise observe. All of our empirical knowledge is based on observation, to be sure. But that does not mean that everything we know can be observed. I know that there are atoms, and that my wife has thoughts, but I do not observe either. For all his shortcomings, Locke was clear on this distinction, as the quotation just cited reveals.

The ideational theory does not depend upon the epistemological standards of science or common sense, which allow indirect knowledge.

24 In a similar vein, Wittgenstein in the *Philosophical Investigations* (1953) assumed that a language in which words referred to sensations or other mental states would be a "private language," by which he meant one that not merely is not but *cannot* be understood by anyone other than the speaker. He then argued that a private language is impossible. See also Malcolm 1954; Castañeda 1967; and Blackburn 1984: 92–109. Cottingham (1997) argues at length that despite popular opinion, Descartes himself was in no way committed to thought being an essentially unknowable process.

If we insist that only what can be observed can be known, we would have to conclude that we cannot know the thoughts of another person. This would not, however, force us to reject ideational theories of meaning. For such a standard would just as surely imply that it is impossible to know what others mean by their words, or what words themselves mean. For meanings – whether we are concerned with what speakers mean or what words mean – are not observable. We cannot see or hear what words mean, or what others mean by them. In other words, skeptics about indirect knowledge would have to deny that others mean anything or have any ideas, but not that meaning consists in a relation between words and ideas.

Bennett reformulates the Rylean argument so that it does not depend on the false assumption that meanings can be perceived by the senses. The key point for Bennett is that meanings can be "learned from passive, unexperimental, hands-off observation" (1971: 5). This does not constitute an objection to the ideational theory that I have formulated unless facts about thoughts and intentions cannot be learned in the same way. Suppose that a man uses the word "plane" on a given occasion. We can generally determine whether he means "airplane" or "planing tool" on the basis of what we observe about him, and we can do so without experiment or specialized scientific expertise. We may note, for example, that the speaker has mentioned airports, 747s, pilots, and flying, but has said nothing about wood, carpentry, or planing. But we can also determine on the same basis whether he is thinking about airplanes or planing tools, and what his intentions were in using the word. Indeed, I cannot know that S means "airplanes" by "planes" on a given occasion unless I know that he is thinking of airplanes. The privacy objection only works by groundlessly insisting on higher epistemological standards for ideas than for meanings.

§21.5 THE SYNONYMY OBJECTION

The next objection we will consider is the Quinean, which on the surface is just that the ideational theory would sanction talk of synonymy.

> The second point [against positing propositions] is that . . . talk of propositions commits one to a general relation of synonymy, and it is not at all obvious that there is any such general relation. (Field 1978: 40)

Despite its familiarity and popularity, this is a very curious claim. Two words are synonymous iff they have the same meaning. If there are no identities of meaning, then no two words are synonymous. If there are

identities of meaning, then some words are synonyms. In either case, the relation of synonymy exists: it either relates or does not relate any two terms. According to ideational theories, e means μ iff e expresses the idea μ. So ideational theories entail that synonymy is the relation of expressing the same idea. If two terms express the same idea, they are synonymous. If they don't, they aren't. That is, whether or not words are related by the relation of synonymy depends on their relations to ideas. Whereas it may be very difficult to decide whether or not any two terms are synonymous (e.g., "brother" and "male sibling"), the existence of the relation of synonymy seems to be unproblematic.[25]

The problem, according to Field, is that it is difficult or impossible to define synonymy (1978: 49–50).[26] This claim is curious, too. First, since not everything can be defined, the existence of a property does not entail its definability. Second, we have given two familiar definitions of synonymy: one in terms of sameness of meaning, the other in terms of sameness of idea expressed. Quine's own objection was that it is not possible to define synonymy purely in terms of verbal behavior (see Quine 1960, 1961, 1972). But there is no reason other than verificationism to accept the behaviorism on which this objection rests, and very good reason to reject it. Field is not a behaviorist. Nevertheless, he finds the existence of synonymy problematic because "it is very hard to formulate any evidential criteria for two words differing in meaning" (1978: 51). But this is an objection to the existence of the synonymy relation, or to the meaningfulness of synonymy talk, only if a verificationist theory of meaning is assumed. Otherwise, the most we could conclude from the difficulty of formulating evidential criteria is that it is difficult to tell whether or not two words are related by the synonymy relation, that is, whether or not they have the same meaning or express the same idea. The existence of such difficulties is no argument against an ideational theory of meaning. Empirical science is normally difficult because our evidence is incomplete, and often seems hopeless because we have no idea how to gather decisive evidence. The case against behaviorism and verificationism is overwhelming, I believe, and has often been made. I shall not repeat the arguments here.

25 One might imagine Quine the metaphysician (1948) insisting that while the predicate "x is synonymous with y" is true of some pairs of objects, that does not entail the existence of the relation of synonymy. But that issue is irrelevant here, since the ideational theory does not sanction or depend on that particular entailment any more than any other linguistic or scientific theory.

26 Contrast Goodman 1952: 73.

Finally, let me express my opinion that the existence of at least some synonyms is incontrovertible. "Bill's accuser" and "The accuser of Bill" are exactly synonymous, and obviously so to anyone who understands English. So are "premise" and "premiss," "two" and "2." "Two" means the same thing in British English that it does in American English. And "Ja" in German means the same thing as "Yes" in English. I grant that no amount of evidence logically entails these hypotheses. But the same goes for the very best scientific claims.

§21.6 THE IDENTITY OBJECTION

The Quinean will not rest content with the definition of synonymy in terms of sameness of meaning for a reason other than verificationism. For he will now demand to know the *identity conditions* for meanings.

> To defend his thesis that numbers are objects, Frege thought it necessary to provide a sense for statements of identity between numbers, and between numbers and other objects. If, now, we take senses to be objects, we must also provide them with identity-conditions. It is somewhat surprising that Frege himself made no attempt to do this. . . . In Quine's phrase, "No entity without identity." (Linsky 1977: 121–2)[27]

Ideational theories do provide identity conditions for meanings, for they entail that two expressions mean the same thing iff they express the same idea. But this will not satisfy the Quinean either, for now he will want us to provide identity conditions for ideas.

As usually conceived, an identity condition for a class of entities is a statement known a priori specifying the conditions under which two entities in that category are identical. The statement will take the form: *x = y iff x and y satisfy condition C.* Thus defined, the principal "no entity without identity" is a consequence of Leibniz's law, which entails that *two objects of any sort are identical iff they have all the same properties.* This condition holds for ideas as well as for anything else: the idea of a brother is identical to the idea of a male sibling iff the two ideas have all the same properties. If we can establish that the former idea occurred to an individual to whom the latter idea did not occur, for example, we have proved that the two ideas are different. If we are given that the idea of a brother has all of the properties of the idea of a male sibling, including the property of being

27 Cf. Hampshire 1939: 22–4; Pap 1957; Quine 1960: 200–9; Alston 1963b: 82–3; Fodor & Katz 1964: 12; Heath 1967: 178; Hacking 1975: 44; Linsky 1983: 29–30, 33–9; Nelson 1992: 32–3, 127–9, 140–1; Schiffer 1994: 304–5.

composed in a certain way of the ideas "male" and "brother," then we must conclude that the two ideas are identical. Leibniz's law provides part of the foundation for a useful principle of inductive inference as well. If we find that the two ideas occur to the same individuals at the same time, are parts of all the same thoughts, are expressed by the same words, and in all other respects appear to have the same properties, then we have good but not conclusive evidence that the two ideas are identical. For the fact that they appear to have the same properties can be explained by the hypothesis that they do have the same properties, which by Leibniz's law entails that they are identical.

The Quinean will presumably reject such an identity condition as trivial, and as too general to be useful. What is wanted, it might be said, is an informative condition that holds for ideas without holding for everything. One identity condition meeting this condition is based on a familiar law governing the part-whole relationship: *two complex ideas are identical iff they have all the same parts in the same relationships.* The idea of a brother is identical to the idea of a male sibling if, and only if, the former has all the components that the latter has, related in the same way. A correlative principle appears to hold given that there are more than two simple ideas: *two ideas are identical iff they are parts of all the same complexes.* These identity conditions will presumably be rejected, too, on the grounds that they are not *sufficiently* informative. The first cannot be used to determine that two ideas are identical without prior knowledge that some simple ideas are identical. The second rule is of no help in this connection, since it cannot be used to determine that two simple ideas are identical without the antecedent knowledge that some complex ideas are.

Two other plausible identity conditions can be formulated for ideas. The first is the generalization of our semantic theorems: *if ideas x and y are expressed by terms e and f, then x = y iff e and f are synonymous.* Even if the reply given in §13.6 to the Mates objection that troubled Quine is accepted, this synonymy criterion of idea identity will also be rejected as insufficiently informative, since synonymy is defined in terms of the identity of the ideas expressed. The second principle relies on the categorization of ideas as generic events: *ideas x and y are identical iff any occurrence of x is an occurrence of y.* The "only if" clause is a consequence of Leibniz's law: if x and y occur to different people at different times, then they have different properties and so cannot be identical. The "if" clause is not a consequence of Leibniz's law but seems to be equally self-evident: if an occurrence of the idea of a brother *is* an occurrence of the idea of a male

sibling, then the idea of a brother must be the idea of a male sibling. Other generic entities have a similar identity condition. For example, if a performance of Beethoven's Ninth is a performance of Beethoven's last symphony, then Beethoven's Ninth must be his last symphony. And if an occurrence of the word "cat" is an occurrence of the word spelled "c-a-t," then the word "cat" must be the word spelled "c-a-t." For that matter, if an occurrence of pain is an occurrence of c-fiber firing, then pain must be c-fiber firing. It will presumably be argued that this identity condition too is insufficiently informative, on the grounds that we cannot decide whether an occurrence of the idea of a male sibling is an occurrence of the idea of a brother without already knowing that the two ideas are identical.

The critical question is, why does the existence of a class of entities entail the existence of identity conditions meeting any standard of informativeness? And why should we have to know such conditions in advance of detailed investigation of the entities in question? None of the critics of the ideational theory has raised these questions, let alone answered them. Quine might have based an answer on behaviorism or verificationism. But those principles have been refuted. Literally interpreted, the principle "No entity without identity" is self-evident. It merely asserts that everything is what it is and is not another thing. That holds for ideas and everything else, in virtue of the laws of logic. But as interpreted by critics of the ideational theory, the Quinean dictum is much stronger. It asserts that for every class of objects, it must be possible to formulate a priori a noncircular definition of the identity relation over that class that meets an epistemological criterion of informativeness. But what reason is there to think that a relation as fundamental as the identity relation could be given any sort of noncircular definition? And what reason is there to think that reality would be so generous to those in pursuit of knowledge as to provide for every class of entity perfectly reliable signs of identity that can be detected without already knowing that any of the entities in question are identical? As far as I can see, there is none.[28]

Critics of the ideational theory and of other theories postulating "intensional" entities usually point admiringly to sets and material objects as entities whose identity conditions are clear. Sets are identical, it is said, iff they have the same members. This is clearly true when we restrict our attention to sets of mathematical and other abstract objects that exist eternally if they exist at all. When we turn to sets of concrete objects, it is not clear what to say. Consider the Julliard String Quartet, which has

28 Cf. Jubien (1996), who argues vigorously against the claim that there must be such criteria.

changed its members over time. Should we say that there has been one set of string players whose membership has changed over time? Or should we say that there has been nothing but a sequence of different sets, with different memberships defining different sets? I see no way of settling this question nonstipulatively. This unclarity in the notion of a set casts no doubt on the existence of sets or on the legitimacy of using the notion in a mathematical theory.

Material objects are identical, it is said, iff they are spatio-temporally continuous. We are all quite confident, I am sure, that a coffee cup on my desk up to moment t will not suddenly appear on the other side of the room immediately after t, and will not go out of existence after t only to reappear on my desk an hour later. In all the years of human experience, physical objects have never been observed to behave that way. Similarly, we all believe that two physical objects cannot occupy the same space at the same time. We have never seen two billiard balls approach each other on an intersecting course, coalesce perfectly at the point of intersection, and then emerge an instant later to continue on their now-diverging paths. But these are pieces of empirical knowledge, which could well have turned out otherwise. Indeed, given the startling discoveries made in the twentieth century, I would not be surprised to hear physicists claim that such things are possible.

Or consider Theseus's famous boat. Over the years its parts were slowly replaced one by one. The old parts were placed in storage, and much later reassembled into a complete boat. We now have two boats traceable in familiar ways to Theseus's original boat. Which one is really Theseus's boat now? It is profoundly unclear how to answer this question. But that does not make the existence of boats the least bit doubtful. Another puzzle: suppose Morris the cat loses his tail at t. We would ordinarily say that Morris survived this change, and that Morris after t (Morris_2) is identical to Morris before t (Morris_1). But is the spatio-temporal continuity condition satisfied here? It is clearly satisfied in the case of Morris′, that part of Morris before t that consisted of all of Morris except his tail. Morris′ also survived the change at t. But then we get the result that $\text{Morris}_1 = \text{Morris}_2$, $\text{Morris}_2 = \text{Morris}'_2$, $\text{Morris}'_2 = \text{Morris}'_1$. Yet surely $\text{Morris}_1 \neq \text{Morris}'_1$, since one is a proper part of the other. But this violates the transitivity of identity. These puzzles call into question the spatio-temporal criterion of identity. They do not cast doubt on the existence of material objects. Note that these questions about identity conditions for material objects create difficulties for referential theories of meaning, according to which synonymy is having the same referent,

if it is assumed that we have to have "satisfactory" identity conditions for material objects.

Nominalists take words and sentences to be unproblematic, and would consider it a major theoretical advance to reduce talk of ideas and thoughts to talk of words and sentences. Focusing on written language, two sentences are identical, we may say, if they consist of the same letters in the same sequence. Letters, in turn, are thought to be individuated by geometric shape. But it is difficult, if not impossible, to specify precisely all of the possible shapes that any given letter may possess. In addition to an unlimited variety of fonts and handwritings, the possible defects in production are endless.[29] Hence we will be unable in many cases to answer the following question: are these two letters tokens of the same type? A very simple example is provided by typewriters on which the very same shape is used for the number "l" and the lowercase letter "l." What then makes two tokens of "Bill's answer was 'l'" tokens of the same sentence type (both about the numeral, say) rather than tokens of two different sentences (one about the numeral and one about the letter)? Not shape or anything else "objective." The writer's intention is surely relevant. Yet identity conditions are no more easily provided for intentions than for any other intentional state. These difficulties in providing identity conditions for sentences cast no doubt on their existence or theoretical utility.

Consider finally the notion of an epileptic seizure. Suppose that a patient has all of the symptoms of a grand mal seizure that varies widely in intensity for an hour, dissipates entirely for a second, then continues for another hour. Has the patient had one epileptic seizure or two? Is there any "right" answer to this question?

The identity conditions of material objects and physical events – at least those conditions meeting Quinean standards of informativeness – are profoundly unclear. Far from making ideas appear to be "creatures of darkness" by comparison, the examples of sets and physical entities confirm the groundlessness of the claim that an "informative" identity condition must be known a priori to hold universally for every class of entity that there is.

29 See Wetzel 1989 and her *Of Types and Tokens* (forthcoming).

22

Priority Objections

One of the most influential objections to ideational theories charges that ideation does not have the independence needed to explain what meaning is in terms of ideation. The objection is that ideas are not conceptually, ontologically, or epistemologically prior to word meanings. Hence the postulation of ideas to explain meanings does not advance our understanding of meaning, because it is ad hoc, regressive, or circular. To explain meaning in terms of ideas is to explain meaning in terms of itself, or something that can only be known on the basis of itself.

The suggestion that, for a person wittingly to use a significant word, phrase, or sentence, there must antecedently or concomitantly occur inside him a momentary something, sometimes called 'the thought that corresponds with the word, phrase or sentence,' leads us to expect that this supposed internal occurrence will be described to us. But when descriptions are proffered, they seem to be descriptions of ghostly doubles of the words, phrases or sentences themselves. (Ryle 1949: 295)

The ideational theorist attempts to account for the significance of utterances by appeal to thoughts. Thoughts are taken to have a significance that simply gets transferred to the utterance. How thoughts come by their significance is not something these theorists spend much time on. . . . As it stands, the ideational theory of meaning is either circular or incomplete. (Avramides 1989: 142)

In [a *dog-legged* theory], words are thought of as reinterpreted into another medium, such as that of Ideas, whose own powers explain the significance words take on. This idea, I shall argue, is destroyed by considerations which are by now quite familiar in modern philosophy. . . . We face a *regress* of interpretations if we need to introduce another medium whose powers explain the powers of any

given medium. And we are in danger of not advancing at all if the powers of elements of the medium to signify things are left unexplained. . . . [I]t is pointless, a mere shuffle, to introduce an element which can be seen to require just the *same kind* of explanation of the original. (Blackburn 1984: 40, 43)

The major objection to the grammarians' theory of meaning is that it does ignore the central problem: language-world relations. As a result, the theory is seriously incomplete. *At best*, it explains the semantic properties of words in terms of primitive concepts. But, what determines the particular nature of each such concept? In virtue of what does the concept TO CATCH concern catching rather than, say, Maggie's fist or the price of eggs? *At bottom, the grammarians do not explain meaning but take it for granted. For, their primitive concepts are nothing but unexplained meanings.* (Devitt & Sterelny 1987: 102)[1]

Another common theme in these objections is that ideational theories are incomplete in some way if they are not regressive or circular. We will address that objection in the next chapter.

§22.1 THE REFLECTION-OR-IGNORANCE OBJECTION

Pitcher developed Ryle's "ghostly double" metaphor, objecting to the conception of propositions according to which they are the meanings of sentences, and composed of ideas, which are the meanings of words.

Are there rules of meaning-combination as there are rules of word-combination – rules of conceptual syntax? If so, what are they? When one tries to discover what they are, he sees either nothing or mere pale reflections of ordinary syntactical rules – and that ought to make us suspicious. (Pitcher 1964: 8)

In fact, we know that conceptual structure differs to a considerable extent from sentence structure, as elementary facts such as those presented in §13.1, §19.3, and §20.1 demonstrate. All of the evidence presented by Chomsky and his followers for the existence of "deep structures" underlying the "surface structures" of natural language sentences constitutes evidence for the independence of conceptual structure. But suppose there

1 See also Geach 1957a: Chapter 23; Black 1972–3: 276; Lewis 1972: 169–70; Rosenberg 1974: 28–34; McGinn 1982: 243–5; 1997: 105; Blackburn 1984: 67; Devitt & Sterelny 1987: 64–5, 102, 124; and Chierchia & McConnell-Ginet 1990: 46–7, 60. Contrast Harman 1968; Vendler 1972: 144; Schiffer 1972: 14–16; 1982: 120; 1987a: 12–13; 1992: 511; Fodor 1975: 119–22; 1990a: 27, 63–4; Bennett 1976: 27–32; Lyons 1977: 329–30; Loar 1981: 1–5, 8–9, 139–41, 203–4, 238–9; Barwise & Perry 1983: 4; Blackburn 1984: 44, 89, 92, 107, 134–40; Devitt & Sterelny 1987: 64–5, 124–8; Quine 1987; Nelson 1992: 33; Yagisawa 1993a: 146. Devitt and Sterelny cite Katz as a "grammarian."

were a language in which sentence syntax was formally identical to conceptual syntax, so that (a) there is a one-to-one correspondence between expressions and ideas (no ambiguity or synonymy), and (b) a sentence is composed of words in the same way that the thought expressed by the sentence is composed of the ideas expressed by the words. It would still be true that for a sentence in such a language to mean "The sky is blue" is for the sentence to express the thought that the sky is blue.

Alston thought that the force of Ryle's objection lay in the fact that ideas are not identifiable or producible independent of words.

> Take a sentence at random, for example, "When in the course of human events, it becomes necessary for one people to . . . ," and utter it with your mind on what you are saying; then, ask yourself whether there was a distinguishable idea in your mind corresponding to each of the meaningful linguistic units of the sentence. Can you discern an idea of 'when,' 'in,' 'course,' 'becomes,' etc., swimming into your ken as each word is pronounced? In the unlikely event that you can, can you recognize the idea that accompanies 'when' as the same idea that puts in an appearance whenever you utter 'when' in that sense? Do you have a firm enough grip on the idea to call it up, or at least know what it would be like to call it up, without the word being present? In other words, is it something that is identifiable and producible apart from the word? Do you ever catch the idea of 'when' appearing when you utter other words – 'until,' 'rheostat,' or 'epigraphy'?
>
> What is disturbing about these questions is not that they have one answer rather than another, but that we do not know how to go about answering them. What are we supposed to look for by way of an idea of 'when'? How can we tell whether we have it in mind or not? Just what am I supposed to try for when I try to call it up out of context? The real difficulty is that we are unable to spot "ideas" as we would have to in order to test the ideational theory. . . .
>
> If we are to have an explication of meaning in terms of ideas, we must be using 'idea' so that the presence of an idea is decidable independent of determining in what senses words are being used. Ideas would have to be introspectively discriminable items in consciousness. Locke was trying to satisfy this requirement when he took 'idea' to mean something like 'sensation or mental image.' (Alston 1964a: 24–5)[2]

We will take Alston's criticisms in more or less reverse order. First, Alston is wrong in saying that ideas would *have to be* introspectively discriminable items of consciousness. As Katz (1966: 178–83) observed, ideas could be treated as purely theoretical entities postulated to account for linguistic and other behavior. The fact that atoms were not observed but rather

2 See also Skinner 1957: 5–7 and Yagisawa 1993a: 146.

"calculated into existence" to explain observations of macroscopic objects did not render atomic theory incapable of explaining the properties of macroscopic objects.[3] One of the fundamental Chomskyan criticisms of behaviorist and verificationist theories was that they "insisted upon certain arbitrary methodological restrictions that make it virtually impossible for scientific knowledge of a nontrivial character to be attained" (Chomsky 1972: ix).

Second, although ideas are very different from images (§19.2, §19.3), they are nonetheless introspectively discriminable items in consciousness. I know introspectively when the idea of a woman occurs to me as opposed to the idea of a frog. That is, I know when I am thinking about a woman and when I am thinking about a frog. Nothing could be clearer.

I am not claiming that thoughts are always introspectible, nor that every feature of thoughts can be known on the basis of introspection alone. The latter claim, at least, is demonstrably false. I cannot tell introspectively the time at which a thought occurs to me (that requires a clock), whether or not it has occurred to me before (that requires memory), how the thought is related to neurophysiology (that requires complex psychophysical data), which words express it in any given language (that requires knowledge of conventions), and so on.[4]

The presence of an idea is not only detectible, it is also "independently decidable." If you uttered the word "woman," that would almost inevitably make me think of a woman. But *I* do not determine that I am thinking about a woman by first determining what you meant by the word, or what the word means in English. I might infer that *you* were thinking of a woman when you uttered the word from the fact that "woman" means woman in English. But that is not how I determine what I am thinking of when you use the word. Consider in this connection the following sentence, and then think for a second or two before reading on.

Bill bought a plane.

3 Contrast Heath's (1967: 178) criticism of theories that treat concepts as entities: "Every theory of this sort is open to weighty and familiar objections on its own account, but the short answer to all of them is that the abstractions they postulate are not observed, but calculated into existence (or rather, subsistence); that there is no other reason, beyond the demands of theory, for believing them to exist." See also Bloomfield 1933: 17: "The only evidence for these mental processes is the linguistic process; they add nothing to the discussion, and only obscure it." This sort of objection is criticized at length by Katz (1964c), Chomsky (1965: Chapter 1; 1972), Chihara & Fodor (1965), Fodor (1968), and many others.

4 Note the indirect use of introspection in §14.2, where evidence for the thesis that thoughts have constituent structure is provided.

Did the word "plane" in this sentence cause you to think of an airplane, a wood plane, a geometric plane, or all three? Did you think of a level piece of land, or Spain? Did the word make you think of Honolulu or a piano? Whatever you thought of, were you thinking of it before you read the word "plane"? You should know the answer to all of these questions. What I meant by the word "plane" should be irrelevant to your determination. If I should tell you that I was using a code in which "plane" means "kilo of cocaine," that should not lead you to change your answers to any of these questions (although once I say "kilo of cocaine," I am sure you will be thinking of cocaine).

The idea of a woman is not wedded to the word "woman." Indeed, I often catch the idea of a woman occurring to me when other words are uttered, such as "femme," "Frau," "man," "pregnant," and "bikini." An idea connected with one word by the conventions of one language is connected with other words in other languages. And any word has various ideas associated with it other than those that it expresses (see Chapter 18). Note too that I can make the idea of a woman occur in someone else without using *any* word: I could do so by drawing a picture of a woman, producing a photograph of a woman, or, better yet, by producing a woman. The familiar "tip of the tongue" phenomenon is one in which the idea of something, usually a person, is occurring to us, without our being able to call to mind the name expressing that idea. The phenomenon almost never occurs with words like "woman," but if it did, the idea of a woman would be occurrent without the word's being present. So the idea of a woman is most definitely identifiable and producible apart from the word "woman."

Ideas and meanings are not completely independent, of course. To use a word like "plane" in a particular sense is to use it to express a particular idea. If I had not been thinking about a geometric plane when I wrote "plane" in the earlier example, then I could not have meant geometric plane by that word. Turning from speakers to hearers, if the idea of an insect scientist rather than the idea of a word scientist occurs to Sharon when she hears "Dr. Strangelove is an etymologist," then she did not understand the sentence. Facts like these, of course, strongly support an ideational theory of meaning.

Yagisawa argued that speaker reference is "parasitic" on term reference, because "[i]n many cases people do not have any particular object in mind in the appropriate psychological sense when uttering a name which refers. When asked whom they have in mind in uttering, say, 'Cicero,' ordinary people would say something like, "Oh well, *Cicero*, you know, a famous

Roman" (1993a: 146). First, the "ordinary" speaker's response in this case indicates that he *does* have a particular object in mind, namely Cicero, the famous Roman orator. Second, Yagisawa's objection is presumably based on the possibility that the ordinary speaker may have no other way of specifying the person he has in mind except by using the very term whose speaker reference was in question. The speaker may not know much about Cicero, and may not know any other names for or definite descriptions of Cicero. But such a de facto limitation on the ordinary speaker's ability to identify or describe his concepts does not prove that the identity or intrinsic nature of an idea is in any way dependent on the meaning of words. Since it would be possible for people to have Cicero in mind even though the name "Cicero" either meant someone else or had never existed, having Cicero in mind cannot be dependent on facts about "Cicero."

Alston asks whether we can catch an idea "swimming into our ken" as the words "when" and "in" are pronounced. If ideas are defined as images, the answer would be "Surely not."[5] Without the identification of ideas with images, Alston suggests that his question is "funny," because we do not know how to answer it. However, an inability to answer a question is no proof that it is not a meaningful question. Furthermore, nothing in ideational theories of meaning or the definition of ideas as thought-parts entails an affirmative answer to Alston's question. It is possible that no part of the thought expressed by a sentence occurs to a reader until the whole sentence has been read. In fact, though, it seems that we can answer Alston's question. There is evidence that each meaningful word does call up an idea as the sentence is read. This is hard to detect when a sentence is read at normal speed. For then everything is happening too fast. The sequence of events can be observed, however, by slowing down the reading of the sentence. This can be done by flashing one word at a time on a screen, and asking the subject to introspect. As an imperfect simulation of this process, read the following:

Bill . . . kicked . . . Sue . . . when . . . Jack . . . kissed . . . Mary.

Here is what happens in my case. As I read "Bill," I think of Bill and wonder what is going to be said about him. When I read "kicked," I think the incomplete thought "Bill kicked" and wonder "kicked what?" When I read "Sue," I think "Bill kicked Sue." When "when" comes along, I think the incomplete thought "Bill kicked Sue when" and wonder

5 Cf. Berkeley 1710: 1.19–1.20 and Warnock 1953: 74–5.

"when what?" And so on. By the time I have finished reading a sentence in this way, a whole series of ideas has occurred to me, each one more complex than the preceding because it incorporates the idea expressed by the most recently read word.[6] When later words in the sentence force me to reinterpret earlier ones ("Jack kissed the wall with his car"), the last thought (my final interpretation of the sentence) is almost totally unrelated to the first. So at least when sentences are read slowly, the ideas expressed by the words do swim into our ken seriatim (if I am at all typical). We would need more evidence, of course, to say for sure what happens when sentences are read quickly.[7]

Schiffer has recently revived the charge that propositions and the like are "ontologically shallow" reflections of sentences. He claims they propositions are "fishy" because of what he calls their "something-from-nothing" feature.

> From a true sentence containing no singular term that refers to an entity of the kind in question, we get a singular term that does refer to an entity of the kind in question. . . . The sentence 'Fido is a dog,' *whether or not it is true,* also yields the singular term 'the property of being a dog', which we are assured of referring to the property of being a dog, and the singular term 'that Fido is a dog', which we are assured of referring to the proposition that Fido is a dog. (Schiffer 1994: 304–5)

First, sentences often generate singular terms referring to objects not mentioned in the sentences. Given the truth of "John has a wife," we know that "John's wife" has a referent, even though the sentence does not contain a singular term referring to John's wife. And whether true or not, the sentence "The sky is blue," which contains no singular term referring to a sentence, yields '"The sky is blue,"' which does refer to a sentence. There is nothing fishy about this process. Second, "the proposition that p" is not guaranteed to have a referent unless the sentence replacing "p" is meaningful. It also has to be declarative rather than interrogative or imperative. We are not getting something from nothing

6 A similar process occurs when speaking. We start with a thought, which may be indefinite, say "Bill kicked someone." As we begin to say "Bill . . . kicked . . . ," we think "Who did Bill kick? Sue." So then we add "Sue . . ." to our sentence. While we are doing that we think "When did Bill do this?" leading us to complete the sentence as indicated, thereby expressing one of the many thoughts we had during the production process. Geach's (1957b: 106) conclusion that "[a]ll we can say is that the judgment is loosely bound up with time" was premature.

7 Indirect evidence that words do call up ideas as they are read is being gathered in experiments on natural language processing. See, e.g., Stillings et al. 1995: Chapter 11.

when we infer the existence of a proposition from the premise that a declarative sentence is meaningful. To be a meaningful declarative sentence is to express a proposition. The convention of using expressions of the form "the proposition that p" to refer to the proposition expressed by "p" when "p" is a meaningful declarative sentence is based on the fact that the meaning rules of a language pair such sentences with propositions (§7.6, §21.2). More fundamentally, such ideo-reflexive reference is possible because we are aware of what we express and can think about and refer to it.

In sum, ideas are not mere reflections of words. While there is much that we do not know about ideas, lack of omniscience about a class of entities is no criticism of theories that postulate them. Besides, there is much that we do know.

§22.2 THE REGRESS OBJECTION

One of the classic "straw men" of semantics is the purely "translational" theory, according to which word e is said to have meaning because a related word f has meaning. Since we want to know what it is for words in general to have meaning, this approach is unilluminating. The translational account is equally useless whether f is in a well-known learned public language or a speculative innate private language.[8] To explain the meaning of one word in terms of the meaning of another begins a regress that gets us nowhere if we are trying to understand what word meaning is.

The passages from Avramides, Blackburn, and Devitt and Sterelny quoted at the beginning of the chapter claim that the ideational theory is a regressive translational theory. It is not. True, the words "significance" and "representative power" can be applied to both words and ideas. But when so applied, they express *two different phenomena*: meaning, and content. In the sense in which sentences have meaning, thoughts and ideas do not. If the sentence "Grass is green" and the thought that grass is green both meant that grass is green in the linguistic sense, they would be synonymous. But it is nonsensical to say that a thought is synonymous with a sentence. Semantic terms like "ambiguous" and "unambiguous" also fail to apply to thoughts and ideas. Understanding an expression involves

8 This is Cresswell's (1985: 56–7) and Nelson's (1992: 179, 246) criticism of Fodor, which is based on Fodor's taking "inner representations" to be expressions (see Chapter 20, this volume). It is a latter-day version of Ryle's famous regress argument (1949: 295–6). See also Bennett 1971: 1; Hacking 1975: 44; Heil 1980: and Johnson-Laird, Herrmann, & Chaffin 1984: 310. Contrast Barwise 1987: 83.

taking it to mean what it means. We cannot speak of understanding thoughts in this sense. When the term is applied to concepts, "understand" means "grasp." Grasping a concept is very different from understanding a word (see §16.7).[9]

Sentences, but not thoughts, are expressions – vehicles of expression that are normally used for the purpose of communication. What sentences mean is ultimately dependent on human volition, and has something to do with what people mean by them. What thoughts "mean" is not at all dependent on human volition; it has nothing to do with what people mean by them, since people do not use thoughts for any purpose. Since no sentence has its meaning by necessity, as part of its intrinsic nature, it makes sense to ask why the sentence "Grass is green" happens to mean that grass is green. But a thought has its content by necessity. The thought that grass is green could not possibly have had the content that snow is white. Since thoughts have their contents as part of their intrinsic nature, it makes no sense to ask why the thought that grass is green happens to have the content that grass is green. The question "In virtue of what is the concept of grass about grass rather than about snow?" is as misguided as "In virtue of what does H_2O contain hydrogen rather than zinc?" Acquiring a concept is not a matter of learning to attach a particular content to a contentless concept. It is a matter of acquiring the ability to think thoughts containing that concept. Thoughts do not "come by" their contents.[10] A fortiori, thoughts do not have their contents because words or anything else with which they are associated have any sort of representational character.

9 There is a sense of meaning that applies to both utterances and thoughts, namely *evidential* meaning or "information content" (see §2.1). S's *thinking* or *believing* "The cat is on the mat" may *indicate* that the cat is on the mat as surely as S's *saying* "The cat is on the mat" (cf. Barwise & Perry 1983: 15). Evidential meaning differs markedly from both word meaning and ideational content, and is irrelevant in the present context. Note especially that doubting, hoping, and fearing that the cat is on the mat have the same ideational content without having the same evidential meaning.

10 Some have rejected the principle that thoughts have their content necessarily on the grounds that (1) to have a certain content is to be related in a certain way to the objects figuring in that content (e.g., the thought that Secretariat is a horse has the content it does because it is related in a certain way to Secretariat); and (2) any such relation is contingent. But premise (1) is mistaken as we are using "content." Having a content is an *intentional* matter, not a relation (§6.3, §15.6). The thought that Pegasus is a horse has a perfectly good content, but is not *related* in any way to Pegasus. And the thought that JFK was assassinated differs in content from the thought that the thirty-fifth president was assassinated, even though both thoughts stand in exactly the same relationships to JFK and the thirty-fifth president, since they are one and the same person.

It is a common enough occurrence for people to identify a sentence without knowing what it means. Hence the mere fact that sentence A means the same thing as "The sky is blue" does not suffice to tell us what A means, unless we happen to know what "The sky is blue" means. The thought that the sky is blue, by contrast, cannot be identified as the thought that the sky is blue without knowing its content. Consequently, the fact that sentence A expresses the thought that the sky is blue does tell us what A means.

Thought content and sentence meaning are related phenomena, to be sure. Indeed, that is a desired implication of ideational theories. But they are not the same phenomenon.[11]

§22.3 THE DEFINITIONAL CIRCULARITY OBJECTION

Some have thought that the ideational theory is circular not because ideational content is the same thing as word meaning, but because content must be defined in terms of meaning.[12] The objection is that ideas have content because words have meaning, not vice versa. Hence we cannot explain what it is for words to have meaning in terms of ideas.

The circularity objection is also unfounded. For the fact that an individual is thinking about something does not logically entail that words exist or have any particular meaning, or that any languages have arisen. Discoverers, inventors, and scientists regularly think about things for which there are not yet any words. There is plenty of behavioral evidence that children can think about their parents, toys, and food before they learn to speak, and that chimpanzees and orangutans can think about bananas even though they may never learn a language. The facts of evolution make it more than reasonable that before any human language arose, languageless primates existed who were similar to modern chimpanzees. It is hard to avoid the conclusion that they too could think about bananas and other things. Since evolution could have stopped with such creatures, it is logically possible for thinking beings to have existed without meaningful words having existed. It follows that thinking about

11 Cf. Ockham (*Summa Logicae I*: 37, 52), who said that words are "conventional signs" and concepts "natural signs," using "sign" very differently than Hobbes did (*De Corpore*, 2.2). See also Ockham, *Ordinatio*: 46; *Quodlibeta*: 47; Boehner 1964: xxxi; Fodor & Lepore 1992: 44–8; and §6.3 here. Contrast Evans 1982: 284–5; Sterelny 1990: Chapter 66; Chisholm 1990; Peacocke 1992: 134–8.

12 For example, Sellars 1958; 1963: xii–xv; 1969; Alston 1963b: 79; Aune 1967b; Quine 1987: 88; Avramides 1989: 78–9; Grice 1989: 142; Yagisawa 1993a: 146; Gauker 1994: 4–5.

something is not logically or even empirically dependent on the existence of meaningful words.

Weaker sorts of dependencies undoubtedly exist, but do not make the ideational theory circular, despite the worries of Field and others.

> My *guess* is that in a typical case, *part of* what makes a symbol in my system of internal representation a symbol that stands for Caesar is that this symbol acquired its role in my system of representation as a result of my acquisition of a name that stands for Caesar in the public language. If something of this sort is true, it would appear to defeat the above approach to a theory of meaning for a public language. (Field 1978: 53)[13]

Let us grant that *I* am able to think about Caesar only because I have "acquired" the English word that means "Caesar." It does not follow that my acquiring an English word meaning "Caesar" is what *makes* an occurrence to me of the idea of Caesar an occurrence of the idea of Caesar. For that same idea has occurred to people who have never acquired any English words, and could have occurred to people who had never acquired any language. Furthermore, I myself could have acquired the concept of Caesar differently, by acquiring some natural language other than English, by looking at pictures, and so on. Finally, if in *every* individual case an idea's being the idea of Caesar resulted from the acquisition of a public language name for Caesar, then we would get the absurd result that a public language would have to contain a name for Caesar before *any* speakers of that language had ever thought about Caesar.

If the very existence or content of ideas depended on the existence of meaningful words, it would be impossible to explain how languages arose in the first place. Human beings have an innate ability to learn languages, but no innate ability to speak any particular language. Actualization of the innate learning ability requires exposure to some particular language. Unless parents are already speaking a language, children will not learn it by the normal maturational process. How, then, did people come to speak the first language? This is a difficult question on any view. No one can completely explain *any* act of invention or creative insight. But if we add that before people spoke the first language, those people were completely devoid of thought, and therefore devoid of belief, reason, and volition, the question becomes more than difficult. It doesn't help to allow that the people had thoughts without content. If an organism without voluntary

13 Cf. Condillac 1746: I.4.25; Searle 1969: 38; Rosenberg 1974: 27; Kaplan 1989: 604, fn. 89; Wettstein 1991: 167; Fitch 1993: 472; Gauker 1994: 4–5; Robins et al. 1997.

control of its vocal cords, and with no conception of the world around it, and no thought of bears, accidentally produced the sounds "There are bears ahead," its "utterance" would not count as the use of a language. The same would be true even if its output were acoustically indistinguishable from that of a sports broadcaster. Other members of its species, moreover, would perceive no connection between its utterances and bears, if they too were unthinking entities with neither a concept of sound, nor the ability to recognize bears or other objects.

By contrast, there is no incoherence in the following hypothesis: a languageless being with thoughts, beliefs, desires, and the capacity for intentional action developed a desire to do something that would indicate the thought that there are bears ahead, and somehow came to believe that making a "grr" sound would do so. Others observed the action and figured out why it was performed. They later copied the action when they had the same desire. The practice spread and evolved, with the result that the different sound "gor" came to mean "There are bears ahead." More words were added, further sound shifts occurred, and word order and other elements of structure became significant. After a long evolutionary process, "There are bears ahead" came to mean what it does.[14]

Peacocke has suggested that thought might be defined in terms of language even though thought is not dependent on language.

> It is indeed highly implausible that all attitudes with content require possession of a language: the spatial reasoning and thought of one of Koehler's apes tells against any such requirement. But there is no contradiction in holding simultaneously that a philosophical explanation of one concept must make use of another and that nevertheless something can fall under the first without falling under the second. No one doubts that what it is for an object to be a piece of currency must be explained by means of the concept of an exchange of goods or services between two parties; but there is no difficulty in the idea of a piece of currency which is never actually exchanged. Perhaps 'possible object of exchange' is in some respect analogous to 'content of a possible object of linguistic expression.' (Peacocke 1986: 112–13)

On the other hand, the hypothesis that money existed before the practice of exchange ever arose is nonsensical in a way that the hypothesis that ideas existed before languages arose is not. Furthermore, whereas we could not coherently suppose that prehistoric humans had the concept

14 Cf. Schiffer 1972: 15–16, 119–30; Dummett 1973: 362–3; Bennett 1976: 206–10; Grice 1982; Kasher 1982: 29; Blackburn 1984: 134–40; Suppes 1986: 113; Avramides 1989: 160–8; and Peacocke 1986: Chapter 8. See also Fodor 1975 and Loar 1981: 216.

of legal tender before they had the concept of trade, there seems to be no incoherence in the hypothesis that prehistoric humans had the concept of thought before they had any concept of linguistic expression. As long as they had introspective access to their thoughts, they should have been able to acquire the concept of thought. So Peacocke's analogy is not completely apt.

Nevertheless, in a weak enough sense of possibility it is surely true that an idea is something that could possibly be expressed in language, and that the idea μ is something that could possibly be expressed by a symbol meaning μ. Even if ideas preceded language, we might still insist that the ideas could have been expressed if there had been a language, and are expressed in possible languages. So let us consider the following equivalence as a definition of content in terms of meaning.

(1) Idea i has content μ iff it is possible for i to be expressed by a word meaning μ.

The left-right conjunct of (1) is a consequence of the expression theory. Ideas can be expressed because people can perform various actions with the intention of producing an indication of the ideas. Ideas can be expressed by words because people can conventionally use the words to express the ideas. In particular, given any idea μ, it is possible for people to use a word w to express it, and for such usage to be conventional. In that case, w means μ.

The right-left conjunct of (1) is false, however. Due to the possibility of verbal slips, and codes, it is possible for speakers to use the English word meaning "entomologist" to express the idea "etymologist." And due to the possibility of ambiguity, it is possible for the idea "shrewish woman" to be conventionally expressed by a word meaning "female fox." It is true that if a word means μ *on a given occasion*, then it must express the idea μ on that occasion (§7.9). But there is no guarantee that the idea μ is the *only* idea that the word expresses. Indeed, double entendre is the use of a word or expression to mean two different things at the same time ("Frank's Nursery: A Growing Concern"). Given that it is possible for a phrase meaning "a growing concern" to express both i and j, (1) does not tell us which one has the content "an expanding concern" and which one has the content "a concern for growing." Indeed, (1) would seem to assign *both* contents to both ideas, even though it is absurd to suppose that one idea has two different contents.

Even if (1) were true, there is reason to think that it would not tell us what it is for an idea to have the content μ. For the meaning of a word seems to be a consequence of the idea that it expresses, not vice

versa. "Grass" means what it does because it expresses a certain idea, whose content is "grass." But that idea does not have the content "grass" because it can be expressed by "grass." If (1) did tell us what it is for an idea to have the content "grass," say, then what account could we give of what it is for a word to mean "grass"? Meaning "grass" is not an intrinsic, essential property of a word in the way that having the content "grass" is an intrinsic, essential property of an idea. So something has to make it true that a word means "grass." The only account that we have seen that is close to being adequate is the thesis that to mean "grass" is to express the idea "grass." So the possibility that Peacocke suggests is not a serious possibility, and does not show that meaning is prior to thought in any way.

Many arguments for the alleged circularity of the Gricean analysis rest on simplistic epistemological assumptions, the verificationist theory that the meaning of a statement consists in its verification conditions, and a failure to distinguish what is possible in practice from what is logically possible.

> The difficulty for Grice is clear: detection of sentence-meaning will require on his programme the detection of utterers' intentions, and, perhaps, of audiences' responses. But again, for any except the simplest of intentions and simplest of responses, such detection will require both the verbal expression of those intentions and responses together with an *understanding* of those verbal expressions. This verbal expression of intention or response, if sufficiently precise, will standardly use, in part, the very sentence whose meaning we are trying to detect *via* the intentions held and responses induced. So to gain access to the appropriate intentions and responses we must first have knowledge of the meaning of the sentence concerned. (Platts 1979: 91)[15]

Let me first note that knowledge of the meaning of a sentence is often of little help in determining what the speaker means by it. Suppose that S says "The vice-presidents saw the head." This sentence has a fantastic number of meanings in English: "vice-president" may mean "U.S. vice-president," "vice-president of any country," or "corporate vice-president"; "saw" may mean "visually perceive," "met," or "cut with a saw"; "head" may mean "head of the body," "head of the department," "head of the beer," "flower cluster," "toilet," and many other things. Moreover, the speaker may have been speaking metaphorically, using a code, or committing

15 See also Davidson 1973: 127, 134; 1974: 143–4; 1975; Wright 1975: 372, 376; Kempson 1975: 141; Biro 1979. Contrast H. H. Clark (1983), who stresses the ubiquity of "nonce sense."

a parapraxis. Second, I reject the assumption that complex intentions *cannot* be identified unless they are expressed verbally. A fighter pilot may be able to tell at a glance that his adversary is trying to avoid being destroyed by heat-seeking Sidewinder missiles by flying into the sun, thereby confusing the missile's guidance system. Hence the pilot may recognize a very complex intention without his adversary's saying a thing.

But let us grant for the sake of argument that detection of intentions requires understanding verbal expressions. The conclusion that an intentionalist analysis is circular still does not follow. First, the verbal expression by which we recognize the intentions and beliefs involved in meaning μ need not be the very expression e whose meaning we are trying to determine. Even if it is, the known meaning of e on the basis of which we recognize the speaker's intentions need not be the meaning that we end up assigning to e. I do not think that it could possibly be maintained, for example, that we somehow discovered that "Time flies" means "Time goes by quickly" *before* we learned that people use "Time flies" to mean that time goes by quickly. A fortiori, we could not have learned that people mean that time goes by quickly on the basis of our knowledge that "Time flies" means "Time goes by quickly." Of course, it is plausible in this case that audiences first figured out what speakers mean by "Time flies" because they knew that "Time flies" means "Time travels through the air on wings." So audiences eventually learned a second meaning of the sentence on the basis of their knowledge of the first. Given such a possibility, an intentionalist analysis might be regressive, but it cannot be circular.

With "He kicked the bucket," the idiomatic meaning bears no evident relation to the literal meaning. Moreover, it is possible to learn what people mean by a sentence without having *any* prior knowledge of what that sentence means. For example, a group of foreigners might point to a person and say "Zoran." They might later point to a group of running antelope and say "gleeting." Later, when they say "Zoran gleeting" in full view of the indicated person running, we might well deduce that by "Zoran gleeting" they meant that that person was running. Here we figured out what people meant by a sentence without *first* knowing what it or any of its components meant. Of course, we might be wrong about what people mean by "Zoran gleeting." But then we might also be wrong about what the expression means, for the same reason. It should also be remembered how easy it often is to figure out what people mean by completely nonlinguistic signals. If a deaf-mute walked into a doctor's office, pointed to a baby doll, pretended to vomit, grabbed the doctor

and gestured to the door, a good guess would be that she meant "Come, my baby is vomiting."[16] That guess would be confirmed if indeed the doctor found the baby doing just that. So the Gricean analysis is not even regressive.

Finally, let us gamely assume, contrary to fact, that it is impossible to learn that people mean μ unless we first know that e means μ. It still does not follow that an intentionalist analysis is circular. For this *epistemological* fact does not entail that the *meaning* or even the *truth conditions* of "People mean μ" somehow include "e means μ." The epistemological fact may be a mere consequence of the limitations of human cognitive faculties. Even if one human being cannot know what another person is thinking without inferring it from the words he utters, there is no logical necessity here. If humans had telepathic powers, divine omniscience, or sufficient knowledge of the neurophysiological correlates of beliefs and intentions, we would not have to rely on linguistic behavior.

§22.4 THE METALINGUISTIC CIRCULARITY OBJECTION

Suppose that we ask why the word "man" means what it does, namely, "man." The ideational theory answers that "man" means "man" because it is conventionally used to express a certain idea, namely, the idea of a man. The fact that our conventional method of referring to the idea of a man is to use the phrase "the idea of a man," which contains the very term whose meaning we are trying to explain, may have led some to think that the ideational theory provides a circular explanation of meaning. Such an objection, however, involves a confusion of what is said in explaining why a word has a meaning, with the conditions for saying it in a particular way. In order for us to say that *i* is the idea of a man by saying "*i* is the idea of a man," the words "is," "the," "idea," and "man" must have certain meanings. But in saying that *i* is the idea of a man, I am not saying anything about words; hence I am not saying that "is," "the," "idea," and "man" have the meanings that they must have for me to have said what I did in the way that I did. It is thus fully possible for someone to say in French that *i* is the idea of a man, without knowing or saying anything about English. Since the *statement* that *i* is the idea of a man says nothing about words, and entails nothing about what any word means, the presence of that statement in an explanation

16 Reid (1764: Chapters 4.2 and 5.3) argued that without a "natural" knowledge of such "natural" signs of our thought, humans would never acquire a language.

of what the English word "man" means will not make the explanation circular.

More generally, *any* explanation *in words* of what it is for words to have a meaning will presuppose the fact that is being explained, namely, that words have meanings. For we cannot use words to say anything unless those words are meaningful. This elementary fact about the nature of explanation does not make it impossible to provide a true and illuminating explanation in words of what it is for words to have meaning. It implies only that, to be true, the explanation provided must also hold for the words being used.

A more sophisticated version of this objection was raised by Lyons in his discussion of "componential analysis," the view that the sense of every meaningful unit of the language can be analyzed as a combination of more basic sense-components. As an example, he says that "the sense of 'man' . . . might be held to combine (in the molecular concept 'man') the atomic concepts 'male', 'adult' and 'human' " (Lyons 1977: 317).

> [W]e can extract from the semantic proportion 'man' : 'woman' :: 'stallion' : 'mare' the factors *a*, *b*, *c*, *d*. . . . The factors would derive their linguistic significance from the fact that each of them enables the linguist to account for the semantic acceptability or unacceptability of sets of sentences: the presence of *a* in "man" would account for the acceptability of 'That man cuts his own hair' and the unacceptability of 'That man cuts her own hair' or 'That man is pregnant'. . . . Most proponents of componential analysis . . . would wish to say that the labels chosen to identify the components have more content than our algebraic factors *a, b, c* and *d*. . . . It is obvious, however, that, unless one provides some extensional definition of MALE, HUMAN, etc., or some intensional definition that does not make metalinguistic use of the English words 'male', 'human', etc. (or the French words 'mâle', 'humain', etc. . . .), no explanation has been given of the meaning of 'male', 'human' or of the sense-components in the meaning of 'man' that adds anything to an analysis that makes use of purely algebraic symbols (cf. Lewis, 1972). (Lyons 1977: 329–30)

To say this is to make the very mistake noted earlier. The information that there are *n* unspecified factors *a, b, c, d, . . .* whose possession in various combinations by the words of various languages explains the linguistic meaning of those words would be a substantive piece of knowledge. But such information does not tell us what the explanatory factors are, and so leaves a major question about meaning unanswered. Without knowing what *a* is, the statement that "male" means "male" because it has factor *a* provides little illumination about why "male" has the meaning that it does, or about what it is for "male" to mean "male." If we are told that

a is the property of expressing the idea of a male, or of denoting the set of males, or of connoting the property of being male, we have been told (correctly or incorrectly) what *a* is. None of these identifications would make the explanation circular, even though our manner of stating these identifications presupposes that "male" means "male," which is the fact that we are trying to explain.

It may be thought that a statement like (2) is at least circular as a definition, on the grounds that the definiendum appears in the definiens in the very sense that we are concerned with.

(2) e means "man" iff e expresses the idea "man."

If successful, this objection would impugn referential theories as well as ideational theories. For statements like the following have the same property.

(3) e means "man" iff e denotes the set of men.
e means "man" iff e connotes the property of being a man.

However, the objection is invalid. It is true that the term "man" appears in each definiens in the sense of the term that we are concerned with. That is, we are trying to explain what it is for a word to mean "man," and we are using the word "man" with that meaning in the definientia of these definitions. It is also true that these definitions cannot be used to explain the meaning of the term "man" to anyone who does not already know its meaning. However, these definitions are not intended to be definitions of the term *man*. They are offered as definitions of the term *meaning "man."* And that term does not appear in any of the definientia. What would be circular are definitions like the following:

(4) x is a man iff x is represented by the idea "man."
x is a man iff x is a member of the set of men.
x is a man iff x has the property of being a man.

While true enough, and illuminating to some extent, such definitions would not tell us what it is to be a man, and would not enable anyone to learn what the word "man" means. In other words, definition (2) and its colleagues cannot be used to explain *what "man" means*, or *what it is to be a man*, for the reason given. But they can be used to explain *what 'meaning "man" ' means*, or *what it is to mean "man."* And that is all that they are intended to explain.

Statements like (2) appear to be circular because of our customary practice of *referring* to ideas by using the words that conventionally

express those ideas (§7.6). We thus use the word-meaning correspondences institutionalized in our language not only to describe objects in the world and communicate our thoughts about them, but also to refer to our ideas. Other methods of referring to ideas are available, although they require empirical knowledge that we may not always possess. All that we need is a contextually unique description of the idea. Suppose, for example, that we have been heatedly discussing the idea of a man, arguing about its proper componential analysis. And suppose that we have finally established that the idea of a man is composed of the ideas "male," "adult," and "human." Then we have two new ways of referring to the idea of a man, which do not use the word "man" or any synonym. We can refer to it as *the idea that we have been trying to analyze*, and as *the idea composed (in a certain way) of the ideas "male," "adult," and "human."* Given that we are introspectively aware of what is on our minds, we could also refer to the idea demonstratively as *that idea*. Such definite and indexical descriptions can in turn be used to fix the reference of names for ideas, such as "idea #5." While we seldom refer to proper thought-parts in this way, the naming of propositions has been a common practice since Euclid (e.g., "Theorem 25"). Then (2) would be equivalent to the following:

(5) e means "man" iff e expresses:
the idea we have been trying to analyze;
the idea of an adult male human;
the idea composed of the ideas "male," "adult," and "human";
that idea;
idea #5.

None of these formulations even appears to be circular. Note finally that being a definition schema, Definition 7.6 does not have the property that (2) has, even though (2) is one of its substitution instances. The symbol "μ" in Definition 7.6 is a placeholder for a word with ideational meaning, not a word with ideational meaning itself. Furthermore, the definitions given for expression (Definition 3.6, Theorem 11.2) do not contain ideo-reflexive expressions, which refer to ideas by using words that express them.

So ideational theories cannot be accused of circularity. Nor is this simply the first step of an infinite regress. In saying that "green" means "green" because it expresses the idea "green," the ideational theorist is not committed to claiming that the idea "green" has the content "green"

because it expresses or is otherwise related to something else that has the content "green." Indeed, we are not committed to saying anything about what it is for an idea to be the idea of green, or for the idea of green to occur to us. If we do, we are likely to say something quite unlike Definition 7.6, citing functions, causal roles, physical correlates, and the like.

23

Incompleteness Objections

We argued in the previous chapter that ideational theories of meaning are neither regressive nor circular. Ideas are logically and epistemologically independent of words. And the intentional content of ideas is a radically different sort of property than the meaning of words. Many have charged that if this is true, then ideational theories of meaning are incomplete. The objection is that ideational theories cannot provide a satisfactory explanation of what it is for a word to have a meaning without explaining what it is to have an idea, or what it is for an idea to be of something. Ideational theories of meaning are also thought to be incomplete because they ignore language-world relations, that is, reference. We will conclude our defense of the expression theory by responding to these objections.

§23.1 UNDEFINED TERMS

We have defined meaning as the expression of ideas and other mental states. We have defined the term "expression" for both speakers and words. Speaker expression was defined in terms of intention and evidential meaning. Word expression was defined in terms of speaker expression and convention. We have defined ideas in terms of thoughts. An idea, we have said, is a thought or a cognitive thought-part (Definition 15.1). We think a thought, or conceive an idea, when it occurs to us. We spent several chapters clarifying the term "thought" to make sure that it was clearly understood, and that we know what it denotes. We all know introspectively what it is to think. We argued at length that thoughts have parts. We did not, however, define "thought." It was our key primitive term.

Every theory has primitive terms and postulates. It is not possible to define, prove, or explain everything. The fact that theories of meaning do not define their primitive terms does not mean that they are *incomplete* in any way. It does prove that they are *limited*. But every theory is limited in the same way. Every theory has undefined terms. No theory of meaning can possibly define all of the terms that it uses to define meaning. It might be said that the terms we have used to define meaning stand as much in need of clarification as does meaning itself. I believe that this is true. But the same can be said for some of the best definitions around, such as "Triangles are three-sided polygons," "A prime number is a number divisible only by one and itself," "Motion is change of place," and so on. A definition cannot tell us what something is if its definiens is obscure or unintelligible. But as long as we understand a term, we can use it to define other terms. As long as we know what it is to think and express an idea, we can use these notions to define meaning. In fact, one way to increase our understanding of a concept is to discover that it can be used to define other concepts. I hope we have increased our understanding of "thought" by showing exactly how it is related to "meaning."

If "thought," "idea," and "content" were linguistic concepts – if they denoted the same sort of thing that "meaning" denotes – then it could be said that ideational theories of meaning were incomplete. In that case, defining "meaning" in terms of "thought" would not enable us to explain what meaning is. But as argued in the previous chapter, the psychological phenomena denoted by our primitive terms are very different from the linguistic phenomena that we have sought to define.

The expression theory does not tell us what makes an event-type a thought, or what it is for a thought to occur to us. The theory does not tell us what it is for the idea of Aristotle to be of Aristotle, nor what it is to conceive that idea. Since these questions are psychological or metaphysical rather than linguistic, they are not the sort of question that a theory of meaning should answer. Nor is there any reason to maintain that the ideational theory is defective until such questions have been answered. Consider the Millian claim that the meaning of the name "Aristotle" is the man Aristotle, or the universally accepted view that the referent of the former is the latter. These theories cannot be rejected or held suspect on the grounds that metaphysical questions like the following have not yet been (and may never be) answered: What makes someone Aristotle? In virtue of what is Aristotle at one time the same individual as Aristotle at an earlier time? There is no more reason to fault the ideational theory because it does not tell us what makes something the idea of Aristotle

than there is to fault the referential theory for not telling us what makes something Aristotle.[1]

Schiffer (1987a: xiv) has queried, "What would be the point of trading in facts about *meaning* for facts about the *content of beliefs* if one ends up with nothing to say about the latter?"[2] The point is that facts about meaning and facts about mental states *are facts*. Moreover, the relationship between these conceptually distinct bodies of fact is not self-evident. Consequently, it would represent an intellectual advance, an increase in our understanding, to discover *further facts* linking the semantic and the psychological. If this seems inadequate, imagine someone asking John Dalton, "What is the point of trading in facts about macroscopic objects for facts about microscopic objects?" The answer would be obvious: the latter *explain* the former. Similarly, the fact that "Time flies" is conventionally used to express the belief or thought whose content is "Time goes by quickly" explains what it is for "Time flies" to mean "Time goes by quickly," and the conditions that led such a convention to arise explain why "Time flies" means that.

§23.2 DEFINING THOUGHT

If anyone should succeed in defining what it is to think, believe, want, or intend something, that would be a stunning intellectual achievement. What are the prospects for success? Decades of philosophical research have made it evident that belief, desire, and thought are fundamental psychological concepts. It does not seem possible to define these propositional attitudes in terms of others, the way it is plausible to say that S is optimistic that p iff S both believes and desires that p. The only alternative is to define the fundamental triad in nonpsychological terms. Given the

1 Cf. Devitt and Sterelny (1987: 124–5), who claim that Grice's failure to specify that in virtue of which intentions and beliefs have the content they do is a "serious weakness" of the Gricean program. Their theory is no more complete. To avoid the circularity problem, for example, they argue that thought content is not universally dependent on conventional word meaning. "*Some* people *once* used 'Socrates' to think about the philosopher without depending on the convention: the people who named him" (125). But one could just as well ask, "In virtue of what were the people who named Socrates thinking about him?" The mere fact that they were looking at or even seeing Socrates does not entail that they were thinking about him. They might have been thinking about his arm, or about the crowd of people he was a member of, to mention just two of the more realistic possibilities. This is Devitt and Sterelny's unsolved "*qua* problem" (1987: §4.4, §5.3). We might also wonder what makes the person the namer saw Socrates.

2 Fodor (1990a: 64) raises the same question about meaning and "use."

success of physics, it is likely that the only ultimate nonpsychological terms available are physical.

Many maintain that any theory invoking mental states with content is somehow suspect until mental states are explicated naturalistically, that is, in the nonintentional terms of physical theory.[3] There is clearly no a priori argument that intentional terms must be so definable, any more than there is that physical terms must be definable in nonphysical terms. I believe there are compelling empirical reasons to think that physicalism is true. A large body of evidence has been successfully explained by a small number of physical laws, and has shown that the properties of macroscopic objects are determined by their microscopic physical parts. Artificial Intelligence ("AI") has been increasingly successful in designing and producing physical objects in which purely physical mechanisms produce good simulations of human behavior. Neuroscience has produced a large body of data connecting the psychological properties of human beings with properties of their neurophysiological parts, specifically those of the brain. We have seen that neural networks provide a good model for ideation (§18.4). So the hypothesis that a thought is a neurophysiological event-type is at least plausible. It does not follow, though, that human beings will ever be able to discover the physical definitions of thoughts or other mental states. There is never a guarantee that scientists will gather the evidence necessary to answer an empirical question.

Even if science is now so advanced that the existence of physical definitions of mental states implies our practical ability to discover them, it would not follow that theoretical use of psychological concepts is prohibited or unsatisfactory *before* physical definitions can be given. Indeed, such a prohibition might prevent the very investigations and discoveries that will eventually produce the physical accounts. Consider as an analogy Dalton's use of atomic theory to explain chemical phenomena. Dalton assumed that the hydrogen atom, for example, was a material object. But he was not able to define hydrogen in physical terms. Physicalism provided no grounds for rejecting Dalton's theory until he was able to give a noncircular physical definition of hydrogen. Such a rejection, moreover, would have precluded the very lines of investigation that led scientists to discover that the hydrogen atom consists of a single electron orbiting a single proton. Physicists and chemists had to learn a lot about hydrogen's behavior as a constituent of various molecules before they could discover

3 See Schiffer 1978: 189–92; 1992: 511; Heil 1980; Maloney 1989: 178; Sterelny 1990: Chapter 6; Cummins 1989: 11; Fodor 1990a: 12–13, 32; 1994: 4–5.

the constituents of hydrogen. In my opinion, both physicalism and the existence of thought and other mental states are sufficiently well established that neither needs to be defended against refutation by the other. We are most likely to discover the physical basis of thought by learning as much as we can about both neurophysiological and psychological phenomena.

Philosophical arguments are often marshaled against the possibility of mental states being neurophysiological. I believe that they were all defused by Place (1956) and Smart (1959). But consider a more recent argument directed against the foundations of the ideational theory of meaning.

> One thing is certain. There is no sense in which elements of a chemical or electrical system intrinsically represent aspects of dogs or carrots. . . . It is quite impossible to see how one element of a chemical or electrical system could in and of itself represent an aspect of a carrot, any more than the fall of a box of paper tissues could in and of itself represent the sinking of the Titanic. (Blackburn 1984: 53)[4]

Here is the counterargument. The idea of a dog represents a dog intrinsically. It is logically possible for the idea of a dog to be a neurophysiological process, and the available evidence indicates that the causes of human behavior are neurophysiological. If a neurophysiological process is the idea of a dog, then it represents a dog intrinsically. Hence it is perfectly conceivable in a logical or epistemological sense for neurophysiological processes to have content intrinsically. Blackburn may have trouble grasping this possibility, but that sort of subjective inconceivability does not refute any scientific theory. The falling tissue box analogy is poor, because what we know about falling tissue boxes makes it completely implausible that they are ideas. We know that boxes represent something only if humans stipulate that they do.

A closer reading will reveal that Blackburn was led astray by some of the key mistakes we have flagged repeatedly.

> The elements of computational mechanisms are just that: elements of a causally complex structure which determines various outputs given various inputs. The elements play a role in this "horizontal" transition from state to state. But they do not also have a God-given "vertical" connection with the world outside the mechanism. . . . (Blackburn 1984: 53)

Blackburn seems to be thinking of "representation" extensionally, as a referencelike *relation* between two things. Neither words, nor neurophysiological processes, nor even ideas, have their extensions intrinsically. But

4 See also Heil 1980; 1981: 330–42; Dennett 1982.

this fact is irrelevant to the ideational theory, which defines the meaning of a word in terms of the *intentional* content of the idea that it expresses (§6.3, §15.4). A neurophysiological process could have Fido as its extension in virtue of two factors: (i) the fact that it is the idea of a Fido together with (ii) the fact that Fido exists, or (i′) the fact that it is the idea of Tom's dog together with (ii′) the fact that Fido is Tom's dog, and so on. Facts of the second type are clearly not inherent properties of neurophysiological processes or ideas.

> Perhaps inside our heads there are structures which because of some fact or other actually do play a representative role for us; perhaps words have their significance because they are mapped into these structures, in some natural and unconscious transformation. This scientific version of a dog-legged theory cannot be ruled out of court. But its immediate appearance is not all that promising. For it shares with the absolute starting-point of the enquiry the need to see the *kind of fact* that is missing: the element of organization, or function, or whatever it might be which makes it true that a given internal structure has a given significance, which it transmits to words which are transformed or mapped into it. But why shouldn't the missing kind of fact which endows it with its representative powers *directly* infuse ordinary straightforward language with its significance? (Blackburn 1984: 54)

Nothing *makes* an idea be the idea of a dog. Hence nothing will *make* a neurophysiological process represent a dog if that process *is* the idea of a dog. And nothing will make it *be* that idea. We need to find evidence that will *show* which neurophysiological processes are ideas. But we do not need to look for the kind of fact that makes them ideas. Since language has its representative powers only because people use and understand it in certain ways, we can be quite certain that what makes a word represent a dog will be completely irrelevant to a neurophysiological process or idea representing a dog.

Neurophysiological definitions might be established on the basis of extensive causal isomorphisms of the sort that functionalists imagine. While functional facts could not be what makes something be the idea of a dog – nothing does – they might help us to determine which neural event is the idea of a dog. Just as the refutation of behaviorism showed behavioral dispositions to be mere data to be explained by psychological hypotheses, so I believe that the many defects of functionalism – which are even more acute for occurrent thought than for belief and desire – show the causal relations of mental states to be limited *clues* that may guide us in discovering their neurophysiological identity. If we find a set of neurophysiological

states that result from observable stimuli, interact with each other, and produce behavior in the same way that a specified set of mental states do *in the cases we have studied*, we would have inductive evidence that the two sets of states are identical. The mental states need not have an absolutely unique causal role for the induction to proceed. The hypothesis would be strengthened considerably by introspective evidence that each mental state in the set occurred when and only when its proposed neurological realization did, especially if scientists learn to produce these states in the laboratory using electrical implants or newer technologies.

Neurophysiological definitions would not make ideational theories circular in any way. Nor is there any reason for anyone interested in semantics to insist that such definitions be produced now, when neurophysiological psychology is in its infancy. Since physicalism is an *empirical* theory, no neurophysiological definitions of psychological properties will be *analytically* or *logically* true. Hence it will always be possible to describe conceivable scenarios in which the definitions fail even if they are true. We can easily *imagine* beings with thoughts but with no bodies at all, or radically different sorts of bodies. Such possibilities do not show that thoughts are not in fact neurophysiological, any more than the fact that we can readily imagine heredity without DNA shows that DNA is not the basis of heredity.

One of the leading arguments against central-state materialism is therefore a fallacy. Stillings and colleagues (1995: 350), for example, reject the theory on the grounds that it is possible for Martians or computers to think without having neurophysiological states.[5] This is surely a logical possibility, which cannot be ruled out on a priori grounds. All that follows, however, is that it is logically possible that thoughts are not neurophysiological. We might plausibly reject the claim that thoughts are neurophysiological events if we knew that it was *empirically* possible for brainless Martians or computers to think – that these were "real" possibilities. But we know nothing of the kind. Logical or linguistic intuition cannot establish empirical possibilities. Indeed, the evidence supporting materialism supports the conclusion that such things are empirically impossible. Moreover, if we did discover thinking Martians without brains, we could simply conclude that thoughts are neurophysiological in humans, and something else in Martians.

Finally, all empirical theories are subject to refutation by future evidence. Indeed, the evidence that mental states are neurophysiological

5 See also Putnam 1967: 200–1; Block 1978: 270, 292. Contrast Kim 1972.

is still far from complete. So if we are not to be unjustifiably dogmatic about an empirical theory, we have to recognize that propositional attitudes might not turn out to be definable in neurophysiological terms either. That possibility need not concern us, since the problems we are addressing in this work can be solved without having an analysis of belief or thought.

§23.3 REFERENTIAL SEMANTICS

Ideational theories of meaning have also been thought to be incomplete because they ignore truth and reference.[6] This is a mistake. Frege's and Russell's problems, among others, show that the meaning or sense of a sentence is not the same as its truth conditions, and that the meaning of a term cannot be equated with its referent. "This is a three-sided polygon" and "This is a three-angled polygon" have the same truth conditions, but different meanings. "Santa Claus" has a meaning but no referent. The fact that a theory attempting to explain what sense is does not define reference does not make the theory incomplete.

If "semantics" is taken to cover reference as well as sense, then a complete semantic theory must include a treatment of reference. Remarkably, most referential theories are incomplete as semantic theories because they do not explain what it is for a word to have a reference. In developing the expression theory, we have defined both speaker reference and word reference. As explained in §7.7, a word e refers to an object x iff e verbally expresses an idea i whose extension is x. The extension of i is x provided that for some "Φ," "i = the idea of Φ" and "Φ = x" are both true. The extension of the idea of Aristotle is Aristotle, and the extension of the idea of the morning star is Venus. Hence "Aristotle" refers to Aristotle, and "the morning star" refers to Venus. "Santa Claus" and the idea that it expresses have no extension, because Santa Claus does not exist. In general, the referential properties of expressions are determined by their meanings together with the facts. On an ideational theory, this means that the referential properties of an expression are determined by the referential properties of the idea that it expresses. The referential properties of an idea are determined by its identity and the facts.

An ideational semantics for a language begins by assigning thoughts and ideas to the sentences and words of the language. The semantics of

6 Cf. Lewis 1969: 171; Cresswell 1985: 27–8, 56–7; Devitt and Sterelny 1987: 28, 33, 102; Chierchia & McConnell-Ginet 1990: 352, 430.

the ideas can be formulated by a Tarski-style recursive theory. Such a theory provides a structural description of ideas, along with assignment of referents to simple ideas, rules assigning referents to complex ideas given the referents of the atomic ideas composing them, rules assigning truth conditions to simple thoughts on the basis of the referents of the ideas composing them, and finally, rules assigning truth conditions to complex thoughts on the basis of the truth-values of the thoughts composing them. The theory can expand its scope by assigning intensions or character functions to the simple ideas and deriving their extensions.

Ideational semantics, and the recursive theory of thought that it is based on, will be developed in my *Nondescriptive Meaning and Reference*.

References

Aaron, R. I. (1967) *The Theory of Universals*, 2nd ed. Oxford: Oxford University Press.

Adams, F. & Fuller, G. (1992) Names, contents, and causes. *Mind and Language*, **7**, 205–21.

Adams, F., Fuller, G. & Stecker, R. (1997) The semantics of fictional names. *Pacific Philosophical Quarterly*, **78**, 128–48.

Adams, F. & Stecker, R. (1994) Vacuous singular terms. *Mind and Language*, **9**, 387–401.

Adams, F., Stecker, R. & Fuller, G. (1993a) Schiffer on modes of presentation. *Analysis*, **53**, 30–4.

Adams, F., Stecker, R. & Fuller, G. (1993b) The Floyd puzzle: Reply to Yagisawa. *Analysis*, **53**, 36–40.

Adams, R. M. (1975) Where do our ideas come from? In *Innate Ideas*, ed. S. P. Stich, pp. 71–87. Berkeley, CA: University of California Press.

Adams, R. M. (1989) Time and thisness. In *Themes from Kaplan*, ed. J. Almog, J. Perry & H. Wettstein, pp. 23–42. Oxford: Oxford University Press.

Algeo, J. (1973) *On Defining the Proper Name*. Gainesville: University of Florida Press.

Allan, K. (1986) *Linguistic Meaning, Vol. I*. New York: Routledge and Kegan Paul.

Allerton, D. J. (1987) The linguistic and sociolinguistic status of proper names. *Journal of Pragmatics*, **11**, 61–92.

Alston, W. P. (1963a) Meaning and use. *Philosophical Quarterly*, **13**, 107–24. Reprinted in *Readings in the Philosophy of Language*, ed. J. Rosenberg & C. Travis, pp. 403–419. Englewood Cliffs, NJ: Prentice-Hall, 1971.

Alston, W. P. (1963b) The quest for meanings. *Mind*, **72**, 79–87.

Alston, W. P. (1964a) *Philosophy of Language*. Englewood Cliffs, NJ: Prentice-Hall.

Alston, W. P. (1964b) Linguistic acts. *American Philosophical Quarterly*, **1**, 138–46.

Alston, W. P. (1965) Expressing. In *Philosophy in America*, ed. M. Black, pp. 15–34. London: George Allen and Unwin.

Alston, W. P. (1967a) Emotive meaning. In *Encyclopedia of Philosophy*, ed. P. Edwards, vol. 3, pp. 486–93. New York: Macmillan.

Alston, W. P. (1967b) Motives and motivation. In *Encyclopedia of Philosophy*, ed. P. Edwards, vol. 5, pp. 399–409. New York: Macmillan.

Alston, W. P. (1967c) Meaning. In *Encyclopedia of Philosophy*, ed. P. Edwards, vol. 5, pp. 233–40. London: Macmillan.
Alston, W. P. (1971) How does one tell whether a word has one, several, or many senses? In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, ed. D. Steinberg & L. Jakobovits, pp. 35–52. Cambridge: Cambridge University Press.
Alston, W. P. (1974) Semantic rules. In *Semantics and Philosophy*, ed. M. K. Munitz & P. Unger, pp. 17–48. New York: New York University Press.
Alston, W. P. (1977) Sentence meaning and illocutionary act potential. *Philosophical Exchange*, **2**, 17–35.
Alston, W. P. (1980) The bridge between semantics and pragmatics. In *The Signifying Animal: The Grammar of Language and Experience*, ed. I. Rauch & G. Carr, pp. 123–34. Bloomington: Indiana University Press.
Alston, W. P. (1982) Review of Holdcroft's *Words and Deeds*. *Noûs*, **16**, 623–6.
Alston, W. P. (1987) Matching illocutionary act types. In *On Being and Saying*, ed. J. J. Thomson, pp. 151–63. Cambridge, MA: MIT Press.
Alston, W. P. (1994) Illocutionary acts and linguistic meaning. In *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, ed. S. L. Tsohatzidis. London: Routledge.
Anderson, J. R. (1983) *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
Anttila, R. (1989) *Historical and Comparative Linguistics*. Philadelphia: John Benjamins.
Aquinas, T. (1268) *A Commentary on Aristotle's De Anima*, trans. R. Pasnau. New Haven, CT: Yale University Press, 1999.
Aquinas, T. (1272) *Treatise on Man*, trans. J. F. Anderson. Englewood Cliffs, NJ: Prentice-Hall, 1962.
Aristotle. *De Anima*. In *The Basic Works of Aristotle*, ed. R. McKeon, pp. 535–606. New York: Random House, 1941.
Aristotle. *De Interpretatione*. In *The Basic Works of Aristotle*, ed. R. McKeon, pp. 40–61. New York: Random House, 1941.
Aristotle. *De Memoria et Reminiscentia*, trans. J. I. Baere. In *The Basic Works of Aristotle*, ed. R. McKeon, pp. 607–17. New York: Random House, 1941.
Armstrong, D. M. (1971) Meaning and communication. *Philosophical Review*, **80**, 427–47.
Arnauld, A. (1641) Objections to the *Meditations*. In *The Philosophical Works of Descartes*, ed. and trans. E. Haldane & G. Ross, vol. 2, pp. 79–95. Cambridge: Cambridge University Press, 1969.
Arnauld, A. (1662) *The Art of Thinking: Port-Royal Logic*, ed. J. Dickoff & P. James. Indianapolis, IN: Bobbs-Merrill, 1964.
Arnauld, A. (1683) *On True and False Ideas*, ed. and trans. S. Gaukroger. Manchester: Manchester University Press.
Arnauld, A. & Lancelot, C. (1660) *General and Rational Grammar: The Port-Royal Grammar*, ed. and trans. J. Rieux & B. E. Rollin. The Hague: Mouton, 1975.
Asher, R. E. (1994) Design features. In *Encyclopedia of Language and Linguistics*, ed. R. E. Asher, pp. 875–7. Oxford: Pergamon Press.
Atlas, J. D. (1989) *Philosophy without Ambiguity*. Oxford: Oxford University Press.
Audi, R. (1982) Believing and affirming. *Mind*, **91**, 115–20.

Augustine. (397) On Christian doctrine. In *Great Books of the Western World, Vol. 18: Augustine*, ed. R. M. Hutchins, trans. J. F. Shaw, pp. 621–98. Chicago: Encyclopedia Britannica, 1952.

Aune, B. (1967a) Thinking. In *Encyclopedia of Philosophy, Vol. 8*, ed. P. Edwards, pp. 101–4. New York: Macmillan.

Aune, B. (1967b) Thinking. In *Knowledge, Mind, and Nature*, pp. 177–211. New York: Random House. Reprinted in *Intentionality, Mind, and Language*, ed. A. Marras, pp. 249–86. Urbana: University of Illinois Press, 1972.

Austin, D. F. (1990) *What's the Meaning of "This"?: A Puzzle about Demonstrative Belief*. Ithaca, NY: Cornell University Press.

Austin, J. L. (1961) The meaning of a word. In his *Philosophical Papers*, pp. 23–43. Oxford: Oxford University Press.

Austin, J. L. (1962) *How to Do Things with Words*. Oxford: Oxford University Press.

Avramides, A. (1989) *Meaning and Mind: An Examination of a Gricean Account of Language*. Cambridge, MA: MIT Press.

Avramides, A. (1997) Intention and convention. In *A Companion to the Philosophy of Language*, ed. B. Hale & C. Wright, pp. 60–86. Oxford: Oxford University Press.

Ayer, A. J. (1952) *Language, Truth, and Logic*. New York: Dover.

Bach, K. (1987a) *Thought and Reference*. Oxford: Oxford University Press.

Bach, K. (1987b) On communicative intentions: A reply to Recanati's "On defining communicative intentions." *Mind and Language*, **2**, 141–54.

Bach, K. (1994a) Conversational impliciture. *Mind and Language*, **9**, 124–62.

Bach, K. (1994b) Meaning, speech acts, and communication. In *Basic Topics in the Philosophy of Language*, ed. R. Harnish, pp. 3–21. Englewood Cliffs, NJ: Prentice-Hall.

Bach, K. (1999) The myth of conventional implicature. *Linguistics and Philosophy*, **22**, 327–66.

Bach, K. & Harnish, R. (1979) *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Bain, A. (1855) *The Senses and the Intellect*. In *Significant Contributions to the History of Psychology*, ed. D. Robinson, vol. 4. Washington, DC: University Publications of America, 1977.

Baldi, P. (1994) Latin. In *Encyclopedia of Languages and Linguistics*, ed. R. E. Asher, pp. 2015–55. Oxford: Pergamon Press.

Barnes, W. H. F. (1944–5) The myth of sense data. *Proceedings of the Aristotelian Society*, **45**, 89–118. Reprinted in *Perceiving, Sensing, and Knowing*, ed. R. J. Schwartz, pp. 138–67. New York: Doubleday Anchor, 1965.

Barwise, J. (1987) Unburdening the language of thought. *Mind and Language*, **2**, 82–96.

Barwise, J. & Perry, J. (1981a) Situations and attitudes. *Journal of Philosophy*, **78**, 668–91.

Barwise, J. & Perry, J. (1981b) Semantic innocence and uncompromising situations. *Midwest Studies in Philosophy*, **6**, 387–404. Reprinted in *The Philosophy of Language*, ed. A. P. Martinich, pp. 392–404. Oxford: Oxford University Press, 1990.

Barwise, J. & Perry, J. (1983) *Situations and Attitudes*. Cambridge, MA: MIT Press.

Bealer, G. (1993a) A solution to Frege's puzzle. In *Philosophical Perspectives, 7: Language and Logic*, ed. J. Tomberlin, pp. 17–60. Atascadero, CA: Ridgeview.

Bealer, G. (1993b) Universals. *The Journal of Philosophy*, **90**(1), 5–32.
Bealer, G. (1998) A theory of concepts and concept possession. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 261–301. Atascadero, CA: Ridgeview.
Bechtel, W. (1988) Connectionism and the philosophy of mind: An overview. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 30–55. Dordrecht: Kluwer. From *Southern Journal of Philosophy,* **26** supplement, 17–42.
Benjafield, J. G. (1992) *Cognition*. Englewood Cliffs, NJ: Prentice-Hall.
Bennett, J. (1971) *Locke, Berkeley, Hume: Central Themes*. Oxford: Oxford University Press.
Bennett, J. (1973) The meaning-nominalist strategy. *Foundations of Language*, **10**, 141–68.
Bennett, J. (1976) *Linguistic Behavior*. Cambridge: Cambridge University Press.
Benson, J. (1967) Emotion and expression. *Philosophical Review*, **76**, 335–57.
Bentham, J. (1816) Universal grammar. In *The Works of Jeremy Bentham, Vol. VIII*, ed. J. Bowring, pp. 184–91. New York: Russell & Russell, 1962.
Bentham, J. (1843) Essay on language. In *The Works of Jeremy Bentham, Vol. VIII*, ed. J. Bowring, pp. 297–338. New York: Russell & Russell, 1962.
Berg, J. (1999) Referential attribution. *Philosophical Studies*, **96**, 73–86.
Bergmann, M. (1982) Metaphorical assertions. *Philosophical Review*, **91**, 229–45. Reprinted in *Meaning and Truth: The Essential Readings in Modern Semantics*, ed. J. L. Garfield & M. Kiteley, pp. 599–612. New York: Paragon House, 1991.
Berkeley, G. (1710) *A Treatise Concerning the Principles of Human Knowledge*, 2nd ed., ed. C. M. Turbayne. Indianapolis, IN: Bobbs-Merrill, 1957.
Bertolet, R. (1987) Speaker reference. *Philosophical Studies*, **52**, 199–226.
Bezuidenhout, A. (1997) The communication of *de re* thoughts. *Noûs*, **31**, 197–225.
Bickerton, D. (1981) *Roots of Language*. Ann Arbor, MI: Karoma Publishers.
Biro, J. (1979) Intentionalism in the theory of meaning. *The Monist*, **62**, 238–58.
Black, M. (1972–3) Meaning and intention: An examination of Grice's views. *New Literary History*, **4**, 257–79.
Blackburn, S. (1984) *Spreading the Word*. Oxford: Clarendon Press.
Blackburn, S. (1992) Theory, observation and drama. *Mind and Language*, **7**, 187–203.
Blakemore, D. (1987) *Semantic Constraints on Relevance*. Oxford: Basil Blackwell.
Blakemore, D. (1992) *Understanding Utterances*. Oxford: Basil Blackwell.
Block, N. (1978) Troubles with functionalism. In *Readings in Philosophy of Psychology*, ed. N. Block, vol. 1, pp. 268–305. Cambridge, MA: Harvard University Press, 1980. From *Perception and Cognition: Issues in the Foundations of Psychology. Minnesota Studies in the Philosophy of Science*, vol. 9 ed. C. W. Savage, pp. 261–325. Minneapolis: University of Minnesota Press: 1978.
Block, N. (1986) Advertisement for a semantics for psychology. In *Midwest Studies in Philosophy Volume 10*, ed. P. French., T. E. Uehling, Jr. & H. K. Wettstein, pp. 615–78. Minneapolis: University of Minnesota Press.
Block, N. (1993) Holism, hyper-analyticity, and hyper-compositionality. *Mind and Language*, **8**, 1–26.
Bloomfield, L. (1933) *Language*. New York: Holt, Rinehart, and Winston.
Boden, M. A. (1991) A horse of a different color? In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich & D. E. Rumelhart, pp. 3–20. Hillsdale, NJ: Erlbaum.

Boehner, P. (1964) Introduction. In *Ockham: Philosophical Works*, ed. P. Boehner. pp. ix–li. Indianapolis, IN: Bobbs-Merrill.

Boër, S. (1986) Chisholm on intentionality, thought, and reference. In *Roderick Chisholm*, ed. R. J. Bogdan, pp. 81–111. Dordrecht: D. Reidel.

Boër, S. (1989) Neo-Fregean thoughts. In *Philosophical Perspectives, 3: Philosophy of Mind and Action Theory*, ed. J. E. Tomberlin, pp. 187–224. Atascadero, CA: Ridgeview.

Boër, S. (1995) Propositional attitudes and compositional semantics. In *Philosophical Perspectives, 9: AI, Connectionism, and Philosophical Psychology*, ed. J. Tomberlin, pp. 341–79. Atascadero, CA: Ridgeview.

Boër, S. & Lycan, W. (1975) Knowing who. *Philosophical Studies*, **28**, 299–344.

Boër, S. & Lycan, W. (1980) Who, me? *Philosophical Review*, **89**, 427–66.

Boër, S. & Lycan, W. (1986) *Knowing Who*. Cambridge, MA: Bradford Books, MIT Press.

Boghossian, P. A. (1993) Does an inferential semantics rest upon a mistake? *Mind and Language*, **8**, 27–40.

Boghossian, P. A. (1998a) What the externalist can know a priori. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 197–211. Atascadero, CA: Ridgeview.

Boghossian, P. A. (1998b) Replies to commentators. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 253–60. Atascadero, CA: Ridgeview.

Bolinger, D. (1980) *Language – the Loaded Weapon*. London: Longman.

Bolinger, D. (1982) Usage and acceptability in language. In *The American Heritage Dictionary*, ed. M. Berube, D. Neely & P. DeVinne, pp. 30–2. Boston: Houghton Mifflin.

Bradshaw, D. E. (1991) Connectionism and the specter of representationalism. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 417–36. Dordrecht: Kluwer.

Brand, M. (1983) Intending and believing. In *Agent, Language, and the Structure of the World: Essays Presented to Hector-Neri Castañeda, with His Replies*, ed. J. E. Tomberlin, pp. 171–93. Atascadero, CA: Ridgeview.

Brand, M. (1984) *Intending and Acting: Toward a Naturalized Action Theory*. Cambridge, MA: MIT Press.

Brandom, R. (1994) *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

Braun, D. (1991) Proper names, cognitive contents, and beliefs. *Philosophical Studies*, **62**, 289–305.

Braun, D. (1993) Empty names. *Noûs*, **27**, 449–69.

Braun, D. (1995) Katz on names without bearers. *Philosophical Review*, **104**, 553–76.

Brentano, F. (1874) *Psychology from an Empirical Standpoint*, ed. L. L. McAlister. New York: Humanities Press, 1973.

Broad, C. D. (1923) The theory of sensa. In his *Scientific Thought*, Chapters 7–8. London: Routledge and Kegan Paul. Reprinted in *Perceiving, Sensing, and Knowing*, ed. R. J. Schwartz, pp. 85–129. New York: Doubleday Anchor, 1965.

Brody, B. (1969) Introduction. In T. Reid, *Essays on the Intellectual Powers of Man* (1785), ed. B. Brody, pp. xii–xxvi. Cambridge, MA: MIT Press.

Bruner, J., Goodnow, J. J. & Austin, G. A. (1956) *A Study of Thinking*. New York: Wiley.

Buckley, W. F. J. (1982) Usage and acceptability in language. In *The American Heritage Dictionary*, ed. M. Berube, D. Neely & P. DeVinne, pp. 32–3. Boston: Houghton Mifflin.

Burge, T. (1975) On knowledge and convention. *Philosophical Review*, **84**, 249–55.

Burge, T. (1977) Belief *de re*. *Journal of Philosophy*, **74**, 338–62.

Burge, T. (1978) Belief and synonymy. *Journal of Philosophy*, **75**, 119–38.

Burge, T. (1979a) Sinning against Frege. *Philosophical Review*, **88**, 398–432.

Burge, T. (1979b) Individualism and the mental. *Midwest Studies in Philosophy*, **4**, 73–121. Reprinted in *The Nature of Mind*, ed. D. M. Rosenthal, pp. 536–67. Oxford: Oxford University Press, 1991.

Burks, A. W. (1949) Icon, index, and symbol. *Philosophy and Phenomenological Research*, **9**, 673–89.

Burks, A. W. (1977) *Cause, Chance, and Reason: An Inquiry into the Nature of Scientific Evidence*. Chicago: University of Chicago Press.

Butchvarov, P. (1996) Conceptualism. In *The Cambridge Dictionary of Philosophy*, ed. R. Audi, p. 148. Cambridge: Cambridge University Press.

Campbell, L. (1994) Language death. In *Encyclopedia of Language and Linguistics*, ed. R. E. Asher, pp. 1960–8. Oxford: Pergamon Press.

Carey, S. (1991) Knowledge acquisition: Enrichment or conceptual change? In *The Epigenesis of Mind: Essays on Biology and Cognition*, ed. S. Carey & R. Gelman, pp. 257–91. New York: Erlbaum. Reprinted in *Concepts: Core Readings*, ed. E. Margolis & S. Laurence, pp. 459–87. Cambridge, MA: MIT Press, 1991.

Carnap, R. (1947) *Meaning and Necessity*. Chicago: University of Chicago Press.

Carnap, R. (1954) On belief sentences. In *Philosophy and Analysis*, ed. M. MacDonald, pp. 128–31. New York: Philosophical Library.

Carr, C. R. (1978a) Expression, meaning, conversation, and indirect speech acts. *Southwestern Journal of Philosophy*, **9**, 89–100.

Carr, C. R. (1978b) Speaker meaning and illocutionary acts. *Philosophical Studies*, **34**, 281–91.

Carruthers, P. (1989) *Tractarian Semantics*. Oxford: Basil Blackwell.

Carruthers, P. (1992) *Human Knowledge and Human Nature*. Oxford: Oxford University Press.

Carston, R. (1994) Thought. In *The Encyclopedia of Language and Linguistics*, ed. R. E. Asher, pp. 4609–10. Oxford: Oxford University Press.

Cartwright, R. (1962) Propositions. In *Analytic Philosophy*, ed. R. Butler, pp. 81–103. Oxford: Basil Blackwell.

Castañeda, H.-N. (1967) Indicators and quasi-indicators. *American Philosophical Quarterly*, **4**, 85–100.

Castañeda, H.-N. (1977) On the philosophical foundations of the theory of communication: I. Reference. *Midwest Studies in Philosophy*, **2**, 165–86.

Castañeda, H.-N. (1985) The semantics and the causal role of proper names. *Philosophy and Phenomenological Research*, **46**, 91–114.

Chastain, C. (1975) Reference and context. In *Language, Mind, and Knowledge*, ed. K. Gunderson, pp. 194–269. Minneapolis: University of Minnesota Press.

Chierchia, G. & McConnell-Ginet, S. (1990) *Meaning and Grammar: An Introduction to Semantics*. Cambridge, MA: MIT Press.

Chihara, C. & Fodor, J. A. (1965) Operationalism and ordinary language. *American Philosophical Quarterly*, **2**, 281–95.
Chisholm, R. M. (1955–6) Sentences about believing. *Proceedings of the Aristotelian Society*, **56**, 125–48.
Chisholm, R. M. (1958) The Chisholm-Sellars correspondence on intentionality. In *Minnesota Studies in the Philosophy of Science*, ed. H. Feigl, M. Scriven & G. Maxwell, vol. 2, pp. 529–39. Minneapolis: University of Minnesota Press. Reprinted in *Intentionality, Mind, and Language*, ed. A. Marras, pp. 214–48. Urbana: University of Illinois Press, 1972.
Chisholm, R. M. (1976) *Person and Object*. La Salle, IL: Open Court.
Chisholm, R. M. (1981) *The First Person*. Minneapolis: University of Minnesota Press.
Chisholm, R. M. (1990) Referring to things that no longer exist. In *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, ed. J. Tomberlin, pp. 545–56. Atascadero, CA: Ridgeview.
Chomsky, N. (1959) A review of B. F. Skinner's *Verbal Behavior*. *Language*, **35**, 26–58. Reprinted in *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 547–78. Englewood Cliffs, NJ: Prentice-Hall, 1964.
Chomsky, N. (1961a) Degrees of grammaticalness. *Word*, **17**, 219–39. Reprinted in *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 384–9. Englewood Cliffs, NJ: Prentice-Hall, 1964.
Chomsky, N. (1961b) On the notion "rule of grammar." *Proceedings of the Twelfth Symposium in Applied Mathematics*, **12**, 6–24. Reprinted in *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 119–36. Englewood Cliffs, NJ: Prentice-Hall, 1964.
Chomsky, N. (1962) A transformational approach to syntax. In *Proceedings of the Third Texas Conference on Problems of Linguistic Analysis of English, 1958*, ed. A. A. Hill, pp. 124–58. Austin: University of Texas Press. Reprinted in *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 211–45. Englewood Cliffs, NJ: Prentice-Hall, 1964.
Chomsky, N. (1964) Current issues in linguistic theory. In *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 50–118. Englewood Cliffs, NJ: Prentice-Hall.
Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press. Chapter 1 reprinted in *Readings in the Philosophy of Language*, ed. J. Rosenberg and C. Travis, pp. 324–64. Englewood Cliffs, NJ: Prentice-Hall, 1971.
Chomsky, N. (1972) *Language and Mind*. New York: Harcourt Brace Jovanovich.
Chomsky, N. (1975) *Reflections on Language*. New York: Pantheon.
Christensen, C. (1997) Meaning things and meaning others. *Philosophy and Phenomenological Research*, **57**, 495–522.
Church, A. (1950) On Carnap's analysis of statements of assertion and belief. *Analysis*, **10**(5), 97–9.
Church, A. (1951) The need for abstract entities in semantics. In *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 437–45. Englewood Cliffs, NJ: Prentice-Hall, 1964. From *Contributions to the Analysis and Synthesis of Knowledge, Proceedings of the American Academy of Arts and Sciences*, **80**, 100–12.
Church, A. (1954) Intensional isomorphism and the identity of belief. *Philosophical Studies*, **5**, 65–73.
Church, A. (1956) Propositions and sentences. In his *The Problem of Universals*, pp. 3–11. Notre Dame, IN: University of Notre Dame Press. Reprinted in *Readings in the*

Philosophy of Language, ed. J. Rosenberg & C. Travis, pp. 276–82. Englewood Cliffs, NJ: Prentice-Hall, 1971.
Churchland, P. M. (1979) *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.
Churchland, P. M. (1980) Plasticity: Conceptual and neuronal. *Behavioral and Brain Sciences*, **3**, 133–4.
Churchland, P. M. (1981) Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, **78**, 67–90. Reprinted in *Mind and Cognition*, ed. W. G. Lycan, pp. 206–23. Oxford: Basil Blackwell, 1990.
Churchland, P. M. & Churchland, P. S. (1983) Stalking the wild epistemic engine. *Noûs*, **17**, 5–18. Reprinted in *Mind and Cognition*, ed. W. G. Lycan, pp. 300–11. Oxford: Basil Blackwell, 1990.
Churchland, P. S. (1980) Language, thought, and information processing. *Noûs*, **14**, 147–70.
Clark, A. (1991) Systematicity, structured representations, and cogntive architecture: A reply to Fodor and Pylyshyn. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 193–217. Dordrecht: Kluwer.
Clark, H. H. (1983) Making sense of nonce sense. In *The Process of Language Understanding*, ed. G. F. D'Arcais & R. J. Javella, pp. 297–331. New York: Wiley. Reprinted in his *Arenas of Language Use*, pp. 305–40. Chicago: University of Chicago Press, 1993.
Clark, H. H. (1993) *Arenas of Language Use*. Chicago: University of Chicago Press.
Clark, H. H. & Marshall, C. (1981) Definite reference and mutual knowledge. In *Elements of Discourse Understanding*, ed. A. Joshi, B. Webber & I. Sag, pp. 10–63. Cambridge: Cambridge University Press.
Cocchiarella, N. B. (1984) Philosophical perspectives on quantification in tense and modal logic. In *Handbook of Philosophical Logic, Vol. II*, ed. D. Gabbay & F. Guenthner, pp. 309–53. Dordrecht: D. Reidel.
Collins, A. M. & Loftus, E. F. (1975) A spreading activation theory of semantic processing. *Psychological Review*, **82**, 407–28.
Condillac, E. B. D. (1746) *Essay on the Origin of Human Knowledge*, ed. R. G. Weyant, trans. T. Nugent. Gainesville, FL: Scholars' Facsimiles and Reprints, 1971.
Cormack, A. & Kempson, R. (1991) On specificity. In *Meaning and Truth: The Essential Readings in Modern Semantics*, ed. J. L. Garfield & M. Kiteley, pp. 546–81. New York: Paragon Press.
Cottingham, J. (1997) Thought and language in Descartes. In *Thought and Language*, ed. J. Preston, pp. 29–49. Cambridge: Cambridge University Press.
Cowie, A. P. (1994) Phraseology. In *The Encyclopedia of Language and Linguistics*, ed. R. Asher, pp. 3168–71. Oxford: Pergamon Press.
Cowie, A. P., Mackin, R. & McCaig, I. (1983) *Oxford Dictionary of Current Idiomatic English*. Oxford: Oxford University Press.
Cowie, F. (1999) *What's Within? Nativism Reconsidered*. Oxford: Oxford University Press.
Crane, T. (1990) The language of thought: No syntax without semantics. *Mind and Language*, **5**, 187–212.
Crane, T. (1991) All the difference in the world. *Philosophical Quarterly*, **41**, 1–25.
Cresswell, M. J. (1985) *Structured Meanings*. Cambridge, MA: MIT Press.

Crimmins, M. (1989) Having ideas and having the concept. *Mind and Language*, **4**, 280–94.
Crimmins, M. (1992) *Talking about Belief*. Cambridge, MA: MIT Press.
Crimmins, M. (1995) Notional specificity. *Mind and Language*, **10**, 464–77.
Crimmins, M. (1998) Hesperus and Phosphorus: Sense, pretense, and reference. *Philosophical Review*, **107**, 1–47.
Crimmins, M. & Perry, J. (1989) The prince and the phone booth: Reporting puzzling beliefs. *Journal of Philosophy*, **86**, 685–711.
Cruse, D. A. (1986) *Lexical Semantics*. Cambridge: Cambridge University Press.
Cummins, R. (1979) Intention, meaning, and truth-conditions. *Philosophical Studies,* **35**, 345–60.
Cummins, R. (1989) *Meaning and Mental Representation*. Cambridge, MA: MIT Press.
Cummins, R. (1991) The role of representation in connectionist explanations of cognitive capacities. In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich & D. E. Rumelhart, pp. 91–114. Hillsdale, NJ: Erlbaum.
Cummins, R. (1996) Systematicity. *Journal of Philosophy*, **93**, 591–614.
Cummins, R. (1997) The LOT of the causal theory of mental content. *Journal of Philosophy*, **94**, 535–42.
Daniels, P. T. & Bright, W., eds. (1996) *The World's Writing Systems*. Oxford: Oxford University Press.
Danto, A. C. (1975) Preface. In *General and Rational Grammar: The Port-Royal Grammar*, ed. J. Rieux, B. E. Rollin & N. Kretzmann, pp. 11–17. The Hague: Mouton.
Davidson, D. (1967) Truth and meaning. *Synthese*, **17**, 304–23. Reprinted in *The Philosophy of Language*, ed. A. P. Martinich, pp. 79–90. Oxford: Oxford University Press, 1990.
Davidson, D. (1968) On saying that. *Synthese*, **19**, 130–46. Reprinted in *The Philosophy of Language*, ed. A. P. Martinich, pp. 337–46. New York: Oxford University Press, 1996.
Davidson, D. (1973) Radical interpretation. *Dialectica*, **27**, 313–28. Reprinted in *Truth and Interpretation*, ed. D. Davidson, pp. 125–39. Oxford: Clarendon Press, 1984.
Davidson, D. (1974) Belief and the basis of meaning. *Synthese*, **27**, 309–23. Reprinted in *The Philosophy of Language*, ed. A. P. Martinich, pp. 456–64. Oxford: Oxford University Press, 1996.
Davidson, D. (1975) Thought and talk. In *Mind and Language*, ed. S. Guttenplan, pp. 7–23. Oxford: Oxford University Press. Reprinted in *The Nature of Mind*, ed. D. M. Rosenthal, pp. 363–71. Oxford: Oxford University Press, 1991.
Davidson, D. (1977) Reality and reference. *Dialectica*, **31**, 247–53. Reprinted in *Inquiries into Truth and Interpretation*, ed. D. Davidson, pp. 215–25. Oxford: Clarendon Press, 1984.
Davidson, D. (1983) Communication and convention. *Journal of Indian Council of Philosophical Research*, **1**, 13–25. Reprinted in his *Truth and Interpretation*, pp. 265–80. Oxford: Clarendon Press, 1984.
Davidson, D. (1986) A nice derangement of epitaphs. In *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, ed. R. Grandy & R. Warner, pp. 157–74. Oxford: Oxford University Press.
Davidson, D. & Harman, G., eds. (1972) *Semantics of Natural Language*. Dordrecht: D. Reidel.

Davies, M. (1981) *Meaning, Quantification, Necessity: Themes in Philosophical Logic*. London: Routledge and Kegan Paul.
Davies, M. (1987) Relevance and mutual knowledge. *Behavioral and Brain Sciences*, **10**, 716–17.
Davies, M. (1991) Concepts, connectionism, and the language of thought. In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich & D. E. Rumelhart, pp. 229–57. Hillsdale, NJ: Erlbaum.
Davis, S. (1994) The Grice program and expression meaning. *Philosophical Studies*, **75**, 293–9.
Davis, W. A. (1981a) A theory of happiness. *American Philosophical Quarterly*, **18**, 111–20.
Davis, W. A. (1981b) Pleasure and happiness. *Philosophical Studies*, **39**, 305–17.
Davis, W. A. (1982) A causal theory of enjoyment. *Mind*, **91**, 240–56.
Davis, W. A. (1984a) A causal theory of intending. *American Philosophical Quarterly*, **21**, 43–54.
Davis, W. A. (1984b) The two senses of desire. *Philosophical Studies*, **45**, 181–95.
Davis, W. A. (1986) *An Introduction to Logic*. Englewood Cliffs, NJ: Prentice-Hall.
Davis, W. A. (1987) The varieties of fear. *Philosophical Studies*, **51**, 287–310.
Davis, W. A. (1988a) Expression of emotion. *American Philosophical Quarterly*, **25**, 279–91.
Davis, W. A. (1988b) A causal theory of experiential fear. *Canadian Journal of Philosophy*, **13**, 459–83.
Davis, W. A. (1992a) Speaker meaning. *Linguistics and Philosophy*, **15**, 223–53.
Davis, W. A. (1992b) Cogitative and cognitive speaker meaning. *Philosophical Studies*, **67**, 71–88.
Davis, W. A. (1998) *Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory*. Cambridge: Cambridge University Press.
Davis, W. A. (1999) Communicating, telling, and informing. *Philosophical Inquiry*, **21**, 21–43.
Davis, W. A. (forthcoming) *Nondescriptive Meaning and Reference*. Oxford: Oxford University Press.
Denkel, A. (1983) The fringes of natural meaning. *Philosophia*, **12**, 337–43.
Dennett, D. C. (1969) *Content and Consciousness*. London: Routledge and Kegan Paul.
Dennett, D. C. (1978) *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: Bradford Books.
Dennett, D. C. (1982) Beyond belief. In *Thought and Object: Essays on Intentionality*, ed. A. Woodfield, pp. 1–95. Oxford: Clarendon Press.
Dennett, D. C. (1986) The logical geography of computational approaches: A view from the east pole. In *The Representation of Knowledge and Belief*, ed. M. Brand & R. M. Harnish, pp. 59–79. Tucson: University of Arizona Press.
Dennett, D. C. (1991) Mother nature versus the walking encyclopedia. In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich & D. E. Rumelhart, pp. 21–30. Hillsdale, NJ: Erlbaum.
Dennett, D. C. (1997) How to do other things with words. In *Thought and Language*, ed. J. Preston, pp. 219–35. Cambridge: Cambridge University Press.
Descartes, R. (1641a) *Meditations*. In *The Philosophical Works of Descartes*, ed. E. Haldane & G. Ross, vol. 1, pp. 131–99. Cambridge: Cambridge University Press, 1969.

Descartes, R. (1641b) *Replies to Objections*. In *The Philosophical Works of Descartes*, ed. & trans. E. Haldane & G. Ross, 2, pp. 1–380. Cambridge: Cambridge University Press, 1969.

Descartes, R. (1644) *The Principles of Philosophy*. In *The Philosophical Works of Descartes*, ed. E. Haldane & G. Ross, vol. 1, pp. 203–302. Cambridge: Cambridge University Press, 1969.

Descartes, R. (1647) Notes directed against a certain programme. In *The Philosophical Works of Descartes*, ed. E. Haldane & G. Ross, pp. 432–50. Cambridge: Cambridge University Press, 1969.

Descartes, R. (c. 1632) *Treatise of Man*, ed. and trans. T. S. Hall. Cambridge, MA: Harvard University Press, 1972.

DeSousa, R. (1971) How to give a piece of your mind: Or the logic of belief and assent. *Review of Metaphysics*, **25**, 52–79.

Deuchar, M. (1984) *British Sign Language*. London: Routledge and Kegan Paul.

Devitt, M. (1981) *Designation*. New York: Columbia University Press.

Devitt, M. (1989) The revival of 'Fido'-Fido. In *Cause, Mind, and Reality*, ed. J. Heil, pp. 73–94. Dordrecht: Kluwer.

Devitt, M. & Sterelny, K. (1987) *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge, MA: MIT Press.

Dickoff, J. & James, P. (1964) Introduction. In *The Port Royal Logic*, pp. xxvii–li. Indianapolis, IN: Bobbs-Merrill.

Donnellan, K. (1966) Reference and definite descriptions. *Philosophical Review*, **75**, 281–304. Reprinted in *The Philosophy of Language*, ed. A. P. Martinich, pp. 235–47. Oxford: Oxford University Press, 1990.

Donnellan, K. (1977) The contingent *a priori* and rigid designators. *Midwest Studies in Philosophy*, **2**, 6–27.

Dorian, N. C. (1981) *Language Death: The Life Cycle of a Scottish Gaelic Dialect*. Philadelphia: University of Pennsylvania Press.

Dressler, W. U. (1996) Language death. In *Towards a Critical Sociolinguistics*, ed. R. Singh, pp. 195–210. Amsterdam: John Benjamins.

Dretske, F. (1969) *Seeing and Knowing*. Chicago: University of Chicago Press.

Dretske, F. (1981) *Knowledge and the Flow of Information*. Oxford: Basil Blackwell.

Dretske, F. (1995) *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Dummett, M. (1967) Frege, Gottlob. In *Encyclopedia of Philosophy*, ed. P. Edwards, vol. 3, pp. 225–37. New York: Macmillan.

Dummett, M. (1973) *Frege: Philosophy of Language*. London: Duckworth.

Dummett, M. (1975) What is a theory of meaning? In *Mind and Meaning*, ed. S. Gutenplan, pp. 97–138. Oxford: Oxford University Press.

Dummett, M. (1976) What is a theory of meaning? (II). In *Truth and Meaning: Essays in Semantics*, ed. G. Evans & J. McDowell, pp. 67–137. Oxford: Clarendon Press.

Dyer, M. (1991) Connectionism versus symbolism in high-level cognition. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 382–416. Dordrecht: Kluwer.

Ebbinghaus, H. (1885) *On Memory*, trans. H. Ruger & C. Bussenius. New York: Teachers College, 1964.

Elugardo, R. (1997) Descriptions, indexicals, and speaker meaning. *Protosociology*, **10**, 155–89.

Emmett, K. (1988) Meaning and mental representation. In *Perspectives on Mind*, ed. H. R. Otto, pp. 77–84. Dordrecht: Kluwer.
Evans, G. (1973) The causal theory of names. *Proceedings of the Aristotelian Society Supplement*, **47**, 187–208. Reprinted in *The Philosophy of Language*, ed. A. P. Martinich, pp. 295–307. Oxford: Oxford University Press, 1990.
Evans, G. (1982) *The Varieties of Reference*. Oxford: Oxford University Press.
Evans, G. (1985) Understanding demonstratives. In his *Collected Papers*, pp. 291–321. Oxford: Oxford University Press. Reprinted in *Demonstratives*, ed. P. Yourgrau, pp. 71–96. Oxford: Oxford University Press, 1990.
Evans, G. & McDowell, J., eds. (1976) *Truth and Meaning: Essays in Semantics*. Oxford: Clarendon Press.
Evans, J. (1989) Concepts and inference. *Mind and Language*, **4**, 29–33.
Evans, J., Newstead, S. E. & Byrne, R. (1993) *Human Reasoning: The Psychology of Deduction*. Hove, UK: Erlbaum.
Fasold, R. (1990) *The Sociolinguistics of Language*. Oxford: Basil Blackwell.
Feit, N. (2000) Self-ascription and belief *de re*. *Philosophical Studies*, **98**, 37–51.
Field, H. (1972) Tarski's theory of truth. *Journal of Philosophy*, **69**, 347–75. Reprinted in *Meaning and Truth: Essential Readings in Modern Semantics*, ed. J. Garfield and M. Kitely, pp. 271–96. New York: Paragon House, 1991.
Field, H. (1978) Mental representation. *Erkenntnis*, **13**, 9–61.
Fine, K. (1989) The problem of *de re* modality. In *Themes from Kaplan*, ed. J. Almog, J. Perry & H. Wettstein, pp. 197–272. Oxford: Oxford University Press.
Fischer, R. (1993) Abbé de l'Epée and the living dictionary. In *Deaf History Unveiled*, ed. J. V. Van Cleve, pp. 12–26. Washington, DC: Gallaudet University Press.
Fitch, G. W. (1985) On the logic of belief. *Noûs*, **19**, 205–28.
Fitch, G. W. (1987) *Naming and Believing*. Dordrecht: D. Reidel.
Fitch, G. W. (1990) Thinking of something. *Noûs*, **24**, 675–96.
Fitch, G. W. (1993) Non-denoting. In *Philosophical Perspectives, 7: Language and Logic*, ed. J. Tomberlin, pp. 461–86. Atascadero, CA: Ridgeview.
Fodor, J. A. (1968) *Psychological Explanation*. New York: Random House.
Fodor, J. A. (1975) *The Language of Thought*. New York: Crowell.
Fodor, J. A. (1981) *Representations*. Cambridge, MA: MIT Press.
Fodor, J. A. (1983) *Modularity of Mind*. Cambridge, MA: MIT Press.
Fodor, J. A. (1987) *Psychosemantics*. Cambridge, MA: MIT Press.
Fodor, J. A. (1989) Substitution arguments and the individuation of beliefs. In *Method, Reason, and Language*, ed. G. Boolos, pp. 63–78. Cambridge: Cambridge University Press. Reprinted in *A Theory of Content*, ed. J. Fodor, pp. 161–76. Cambridge, MA: MIT Press, 1990.
Fodor, J. A. (1990a) *A Theory of Content*. Cambridge, MA: MIT Press.
Fodor, J. A. (1990b) Psychosemantics or: Where do truth conditions come from? In *Mind and Cognition*, ed. W. Lycan, pp. 312–37. Oxford: Basil Blackwell.
Fodor, J. A. (1991) Information and representation. In *Information, Language, and Cognition, 1990*, ed. P. Hanson, pp. 175–97. Oxford: Oxford University Press. Reprinted in *Concepts: Core Readings*, ed. E. Margolis & S. Laurence, pp. 513–524. Cambridge, MA: MIT Press, 1999.
Fodor, J. A. (1994) *The Elm and the Expert: Mentalese and Its Semantics*. Cambridge, MA: MIT Press.

Fodor, J. A. (1998a) *Concepts*. Oxford: Oxford University Press.
Fodor, J. A. (1998b) There are no recognitional concepts; not even RED. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 1–14. Atascadero, CA: Ridgeview.
Fodor, J. A., Garrett, M., Walker, E. & Parkes, C. (1980) Against definitions. *Cognition*, **8**, 263–367. Reprinted in *Concepts: Core Readings*, ed. E. Margolis and S. Laurence, pp. 491–512. Cambridge, MA: MIT Press, 1999.
Fodor, J. A. & Katz, J. J., eds. (1964) *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice-Hall.
Fodor, J. A. & Lepore, E. (1992) *Holism: A Shopper's Guide*. Oxford: Basil Blackwell.
Fodor, J. A. & McLaughlin, B. (1990) Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. In *Connectionism*, ed. C. Macdonald & G. Macdonald, pp. 199–222. Oxford: Basil Blackwell. From *Cognition*, **35**, 183–204.
Fodor, J. A. & Pylyshyn, Z. (1988) Connectionism and cognitive architecture: A critical analysis. In *Connectionism*, ed. C. Macdonald & G. Macdonald, pp. 90–163. Oxford: Basil Blackwell, 1995. From *Cognition*, **28**, 3–71.
Fodor, J. D. (1977) *Semantics: Theories of Meaning in Generative Grammar*. New York: Crowell.
Follett, W. (1970) *Modern American Usage*. New York: Grosset and Dunlop.
Forbes, G. (1987) Indexicals and intensionality: A Fregean perspective. *Philosophical Review*, **96**, 3–31.
Forbes, G. (1989) Indexicals. In *Handbook of Philosophical Logic*, ed. D. Gabbay, pp. 463–90. Dordrecht: Kluwer.
Forbes, G. (1990) The indispensability of *sinn*. *The Philosophical Review*, **99**, 535–63.
Forbes, G. (1993) Solving the iteration problem. *Linguistics and Philosophy*, **16**, 311–30.
Forbes, G. (1997a) Essentialism. In *A Companion to the Philosophy of Language*, ed. B. Hale & C. Wright, pp. 515–33. Oxford: Basil Blackwell.
Forbes, G. (1997b) How much substitutivity? *Analysis*, **57**, 109–13.
Forbes, G. (1997c) Belief reports and speech reports. In *Direct Reference, Indexicality, and Propositional Attitudes*, ed. W. Kunne, A. Newen & M. Anduschus, pp. 313–30. Stanford, CA: CSLI Publications.
Forster, P. G. (1982) *The Esperanto Movement*. The Hague: Mouton.
Frances, B. (1998) Defending Millian theories. *Mind*, **107**, 703–27.
Fraser, A. C. (1894) "Prolegomena" to Locke's *Essay on Human Understanding*, pp. xi–cxl. New York: Dover, 1959.
Freddoso, A. J., ed. (1980) *Ockham's Theory of Propositions: Part II of the Summa Logicae*, trans. A. J. Freddosso & H. Schuurman. Notre Dame, IN: University of Notre Dame Press.
Frege, G. (1879) *Begriffshrift*. In *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach & M. Black, pp. 1–20. Oxford: Basil Blackwell, 1952.
Frege, G. (1884) *The Foundations of Arithmetic*, 2nd ed., ed. and trans. J. L. Austin. Oxford: Basil Blackwell, 1959.
Frege, G. (1892a) On concept and object. In *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach & M. Black, pp. 42–55. Oxford: Basil Blackwell, 1952. From *Vierteljahrsschrift für wissenschaftliche Philosophie*, **16**, 192–205.

Frege, G. (1892b) On sense and reference. In *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach & M. Black, pp. 56–78. Oxford: Basil Blackwell, 1952. From *Zeitschrift für Philosophie und philosophische Kritik*, **100**, 25–50.

Frege, G. (1894) Review of Husserl's *Philosophie der arithmetic*. In *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. T. Geach & M. Black, pp. 79–85. Oxford: Basil Blackwell, 1952. From *Zeitschrift für Philosophie und philosophische Kritik*, **103**, 313–32.

Frege, G. (1918) Thoughts. In *Logical Investigations*, ed. P. T. Geach, pp. 1–30. Oxford: Basil Blackwell, 1977.

Frege, G. (1919) Notes for L. Darmstädter. In *Gottlob Frege: Posthumous Works*, ed. H. Hermes, F. Kambartel & F. Kaulbach, pp. 253–7. Chicago: University of Chicago Press, 1979.

Frege, G. (1923) Compound thoughts. In *Logical Investigations*, ed. P. T. Geach, pp. 55–78. Oxford: Basil Blackwell, 1977.

Frege, G. (1979) *Posthumous Works*, ed. H. Hermes, F. Kambartel & F. Kaulbach. Chicago: University of Chicago Press.

Fries, C. (1940) *American English Grammar: The Grammatical Structure of Present-Day American English with Especial Reference to Social Differences or Class Dialects*. New York: Appleton-Century-Crofts.

Gale, R. M. (1967) Indexical signs, egocentric particulars, and token-reflexive words. In *The Encyclopedia of Philosophy, Vol. 4*, ed. P. Edwards, pp. 151–4. New York: Macmillan.

Garson, J. W. (1991) What connectionists cannot do. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 113–42. Dordrecht: Kluwer.

Gassendi, P. (1641) Objections to the *Meditations*. In *The Philosophical Works of Descartes*, ed. and trans. E. Haldane & G. Ross, vol. 2, pp. 135–203. Cambridge: Cambridge University Press, 1969.

Gauker, C. (1994) *Thinking Out Loud: An Essay on the Relation between Thought and Language*. Princeton, NJ: Princeton University Press.

Geach, P. T. (1957a) *Mental Acts*. London: Routledge and Kegan Paul.

Geach, P. T. (1957b) On beliefs about oneself. *Analysis*, **18**, 23–4.

Geach, P. T. (1969) *God and the Soul*. London: Macmillan.

Geach, P. T. (1980) Some problems about the sense and reference of proper names. *Canadian Journal of Philosophy Supplement*, **6**, 83–96.

Gilbert, M. (1996) *Living Together: Rationality, Sociality, and Obligation*. Lanham, MD: Rowman and Littlefield.

Gillett, G. (1992) *Representation, Meaning, and Thought*. Oxford: Clarendon Press.

Ginnane, W. J. (1960) Thoughts. *Mind*, **49**, 372–90.

Goldman, A. (1970) *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice-Hall.

Goldman, A. (1986) *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldstein, I. (1986) Must there be indefinable words? *Metaphilosophy*, **17**, 90–1.

Goodman, N. (1952) On likeness of meaning. In *Semantics and the Philosophy of Language*, ed. L. Linsky, pp. 67–76. Urbana: University of Illinois Press.

Goodman, N. (1955) *Fact, Fiction, and Forecast*. Indianapolis, IN: Bobbs-Merrill, 2nd ed. 1965.

Goschke, T. & Koppelberg, D. (1991) The concept of representation and the representation of concepts in connectionist models. In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich & D. E. Rumelhart, pp. 129–62. Hillsdale, NJ: Erlbaum.

Graham, G. (1988) Connectionism in Pavlovian harness. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 143–65. Dordrecht: Kluwer, 1991. From *The Southern Journal of Philosophy*, **26**, 73–91.

Grandy, R. (1977) Review of *Convention* by David Lewis. *Journal of Philosophy*, **74**, 129–39.

Grandy, R. (1990) Understanding and the principle of compositionality. *Philosophical Perspectives*, **4**, 557–72.

Grandy, R. (1998) Recognitional concepts and compositionality. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 20–5. Atascadero, CA: Ridgeview.

Grandy, R. & Warner, R., eds. (1986) *Philosophical Grounds of Rationality: Intentions, Categories, Ends*. Oxford: Oxford University Press.

Green, M. (1997) On the autonomy of linguistic meaning. *Mind*, **106**, 217–44.

Green, O. H. (1968) Intentions and speech acts. *Analysis*, **29**, 109–12.

Green, O. H. (1970) The expression of emotion. *Mind*, **79**, 551–68.

Grice, H. P. (1957) Meaning. *Philosophical Review*, **66**, 377–88.

Grice, H. P. (1961) The causal theory of perception. *Proceedings of the Aristotelian Society, Supplementary Volume*, **35**, 121–52.

Grice, H. P. (1967) Indicative conditionals. In his *Studies in the Way of Words*, pp. 58–85. Cambridge, MA: Harvard University Press, 1989. A revised version of part the William James Lectures entitled "Logic and Conversation" presented at Harvard University in 1967.

Grice, H. P. (1968) Utterer's meaning, sentence meaning, and word meaning. *Foundations of Language*, **4**, 225–42. Reprinted in *Studies in the Way of Words*, ed. H. P. Grice, pp. 117–37. Cambridge, MA: Harvard University Press, 1989.

Grice, H. P. (1969) Utterer's meaning and intentions. *Philosophical Review*, **78**, 147–77. Reprinted in his *Studies in the Way of Words*, pp. 86–116. Cambridge, MA: Harvard University Press, 1989.

Grice, H. P. (1975) Logic and conversation. In *Syntax and Semantics, 3: Speech Acts*, ed. P. Cole & J. Morgan. New York: Academic Press. Reprinted in his *Studies in the Way of Words*, pp. 22–40. Cambridge, MA: Harvard University Press, 1989.

Grice, H. P. (1982) Meaning revisited. In *Mutual Knowledge*, ed. N. Smith, pp. 223–43. New York: Academic Press.

Grice, H. P. (1986) Reply to Richards. In *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, ed. R. Grandy & R. Warner, pp. 45–106. Oxford: Clarendon Press.

Grice, H. P. (1989) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Guttenplan, S., ed. (1975) *Mind and Language*. Oxford: Oxford University Press.

Hacking, I. (1975) *Why Does Language Matter to Philosophy?* Cambridge: Cambridge University Press.

Hale, B. & Wright, C. (1997) *A Companion to the Philosophy of Language*. Oxford: Basil Blackwell.

Hall, R. A., Jr. (1962) Review of *Webster's Third New International Dictionary*. *The Quarterly Journal of Speech*, **68**, 434–5.
Halle, M. (1962) Phonology in generative grammar. In *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 334–52. Englewood Cliffs, NJ: Prentice-Hall, 1964. From *Word*, **18**, 54–72.
Hamlyn, D. W. (1967a) Analytic and synthetic statements. In *Encyclopedia of Philosophy*, ed. P. Edwards, vol. 1, pp. 105–9. New York: Macmillan.
Hamlyn, D. W. (1967b) A priori and a posteriori. In *The Encyclopedia of Philosophy*, ed. P. Edwards, vol. 2, pp. 140–4. New York: Macmillan.
Hamlyn, D. W. (1971) Epistemology and conceptual development. In *Cognitive Development and Epistemology*, ed. T. Mischel, pp. 3–24. New York: Academic Press.
Hampshire, S. (1939–40) Ideas, propositions, and signs. *Proceedings of the Aristotelian Society*, **40**, 1–26.
Hampshire, S. (1972) Feeling and expression. In *Freedom of Mind and Other Essays*, pp. 143–60. Oxford: Oxford University Press.
Harman, G. (1968) Three levels of meaning. *Journal of Philosophy*, **65**, 590–602.
Harman, G. (1972) Deep structure as logical form. In *Semantics of Natural Language*, ed. D. Davidson & G. Harman, pp. 25–47. Dordrecht: Reidel.
Harman, G. (1973) *Thought*. Princeton, NJ: Princeton University Press.
Harman, G. (1974) Review of *Meaning* by S. Schiffer. *Journal of Philosophy*, **71**, 224–9.
Harman, G. (1975) Language, thought, and communication. In *Language, Mind, and Knowledge*, ed. K. Gunderson, pp. 270–98. Minneapolis: University of Minnesota Press.
Harman, G. (1977a) Review of Bennett. *Language*, **53**, 417–24.
Harman, G. (1977b) How to use propositions. *American Philosophical Quarterly*, **14**, 173–6.
Harman, G. (1978) Is there mental representation? In *Minnesota Studies in the Philosophy of Science, Volume IX: Perception and Cognition: Issues in the Foundations of Psychology*, ed. C. W. Savage, pp. 57–65. Minneapolis: University of Minnesota Press.
Harnish, R. (1976) Logical form and implicature. In *An Integrated Theory of Linguistic Ability*, ed. T. G. Bever, J. J. Katz & T. Langedoen, pp. 313–92. New York: Crowell. Reprinted in *Pragmatics: A Reader*, ed. S. Davis, pp. 316–64. Oxford: Oxford University Press, 1991.
Harris, R. (1980) *The Language Makers*. Ithaca, NY: Cornell University Press.
Harris, R. & Taylor, T. (1997) *Landmarks in Linguistic Thought: The Western Tradition from Socrates to Saussure*. London: Routledge.
Harrison, G. (1980) *Introduction to the Philosophy of Language*. New York: St. Martin's Press.
Hartley, D. (1749) *Observations on Man*, ed. T. L. Huguelet. Gainesville, FL: Scholars' Facsimiles and Reprints, 1966.
Hartnack, J. (1972) On thinking. *Mind*, **81**, 543–52.
Hatfield, G. (1991) Representation in perception and cognition: Connectionist affordances. In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich & D. E. Rumelhart, pp. 163–95. Hillsdale, NJ: Erlbaum.
Heal, J. (1997) Indexical predicates and their uses. *Mind*, **106**, 619–40.
Heath, P. L. (1967) Concepts. In *Encyclopedia of Philosophy*, ed. P. Edwards, vol. 2, pp. 177–80. New York: Macmillan.

Hebb, D. O. (1966) *A Textbook of Psychology*. Philadelphia: Saunders.
Heidelberger, H. (1982) What is it to understand a sentence that contains an indexical? *Philosophy and Phenomenological Research*, **43**, 21–34.
Heil, J. (1980) Computation, cognition, and representation. *Behavioral and Brain Sciences*, **3**, 139.
Heil, J. (1981) Does cognitive psychology rest on a mistake? *Mind*, **90**, 321–42.
Higginbotham, J. (1998) Conceptual competence. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 149–62. Atascadero, CA: Ridgeview.
Hill, C. O. (1991) *Word and Object in Husserl, Frege, and Russell: The Roots of Twentieth-Century Philosophy*. Athens, OH: Ohio University Press.
Hintikka, J. (1963) The modes of modality. *Acta Philosophica Fennica*, **16**, 65–79. Reprinted in *The Possible and the Actual*, ed. M. J. Loux, pp. 65–79. Ithaca, NY: Cornell University Press, 1979.
Hirschberg, J. (1991) *A Theory of Scalar Implicature*. New York: Garland.
Hobbes, T. (1651) *Leviathan*, ed. A. D. Lindsay. New York: Dutton.
Hobbes, T. (1655) *Computation or Logic*, trans. A. P. Martinich. In *Thomas Hobbes: Part I of De Corpore*, ed. I. C. Hungerland & G. R. Vick, pp. 171–331. New York: Abaris, 1981.
Hockett, C. F. (1958) *A Course in Modern Linguistics*. New York: Macmillan.
Hockett, C. F. (1966) The problems of universals in language. In *Universals of Language*, 2nd ed., ed. J. H. Greenberg, pp. 1–29. Cambridge, MA: MIT Press.
Holdcroft, D. (1978) *Words and Deeds: Problems in the Theory of Speech Acts*. Oxford: Clarendon Press.
Horgan, T. (1998) Recognitional concepts and the compositionality of concept possession. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 28–33. Atascadero, CA: Ridgeview.
Horgan, T. & Tienson, J. (1988) Settling into a new paradigm. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 241–60. Dordrecht: Kluwer, 1991a. From *The Southern Journal of Philosophy*, **26** supplement, 97–113.
Horgan, T. & Tienson, J., eds. (1991b) *Connectionism and the Philosophy of Mind*. Dordrecht: Kluwer.
Hornstein, N. (1984) *Logic as Grammar*. Cambridge, MA: MIT Press.
Horwich, P. (1997) The composition of meanings. *Philosophical Review*, **106**, 503–32.
Horwich, P. (1998a) *Meaning*. Oxford: Oxford University Press.
Horwich, P. (1998b) Concept constitution. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 15–19. Atascadero, CA: Ridgeview.
Horwich, P. (1998c) The deflationary view of reference. *Lingua e Stile*, **33**, 367–81.
Hugly, P. & Sayward, C. (1979) A problem about conversational implicature. *Linguistics and Philosophy*, **3**, 19–25.
Hugly, P. & Sayward, C. (1995) What's so special about sentences? *Communication and Cognition*, **28**, 409–26.
Hull, C. L. (1920) Quantitative aspects of the evolution of concepts. *Psychological Monographs*, **28**, 1–86.
Hull, C. L. (1943) *The Principles of Behavior*. New York: Appleton-Century.
Hume, D. (1739) *A Treatise of Human Nature*, ed. L. A. Selby-Bigge. Oxford: Clarendon Press, 1888.

Humphrey, G. (1951) *Thinking: An Introduction to Experimental Psychology*. London: Methuen.
Hungerland, I. C. (1960) Contextual implication. *Inquiry*, **3**, 211–58.
Hungerland, I. C. & Vick, G. R. (1981) Hobbes's theory of language, speech, and reasoning. In *Thomas Hobbes: Part I of De Corpore*, ed. I. C. Hungerland & G. R. Vick, pp. 1–70. New York: Abaris.
Hunt, E. B. (1962) *Concept Learning: An Information Processing Problem*. New York: Wiley.
Husserl, E. (1891) *Philosophie der Arithmetik*. Halle: Pfeffer.
Husserl, E. (1900) *Logical Investigations*, trans. J. N. Findlay. New York: Humanities Press, 1970.
Hyslop, A. (1977) Grice without an audience. *Analysis*, **37**, 67–9.
Jackendoff, R. (1983) *Semantics and Cognition*. Cambridge, MA: MIT Press.
Jackendoff, R. (1989) What is a concept, that a person may grasp it? *Mind and Language*, **4**, 68–101.
Jacob, P. (1997) Frege's puzzle and belief ascription. In *Direct Reference, Indexicality, and Propositional Attitudes*, ed. W. Kunne, A. Newen & M. Anduschus, pp. 215–45. Stanford, CA: CSLI Publications.
Jacob, P. (1998) Conceptual competence and inadequate conceptions. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 169–73. Atascadero, CA: Ridgeview.
James, W. (1890) *The Principles of Psychology*. New York: Dover, 1950.
Jamieson, D. (1975) David Lewis on convention. *Canadian Journal of Philosophy*, **5**, 73–81.
Janton, P. (1993) *Esperanto: Language, Literature, and Community*, trans. H. Tonkin, J. Edwards & K. Johnson-Weiner. Albany: State University of New York Press.
Johnson-Laird, P. N., Hermann, D. J. & Chaffin, R. (1984) Only connections: A critique of semantic networks. *Psychological Bulletin*, **96**, 292–315.
Johnson-Laird, P. N. & Steedman, M. (1978) The psychology of syllogisms. *Cognitive Psychology*, **10**, 64–99.
Jordan, D. K. (1992) What is colloquial Esperanto? *Esperantic Studies*, **2**, 3–4.
Joseph, H. W. B. (1916) *An Introduction to Logic*, 2nd ed. Oxford: Clarendon Press.
Jubien, M. (1996) The myth of identity conditions. In *Philosophical Perspectives, 10: Metaphysics*, ed. J. Tomberlin, pp. 343–56. Atascadero, CA: Ridgeview.
Kamp, H. & Partee, B. (1995) Prototype theory and compositionality. *Cognition*, **57**, 129–91.
Kant, I. (1787) *Critique of Pure Reason*, trans. N. K. Smith. New York: St. Martin's Press, 1965.
Kapitan, T. (1994) Exports and imports: Anaphora in attitudinal ascriptions. In *Philosophical Perspectives, 8: Logic and Language*, ed. J. Tomberlin, pp. 273–92. Atascadero, CA: Ridgeview.
Kaplan, D. (1969) Quantifying in. In *Words and Objections: Essays on the Work of W. V. Quine*, ed. D. Davidson & J. Hintikka, pp. 206–42. Dordrecht: Reidel.
Kaplan, D. (1977) Demonstratives. In *Themes from Kaplan*, ed. J. Almog, J. Perry & H. Wettstein, pp. 481–563. Oxford: Oxford University Press, 1989.
Kaplan, D. (1989) Afterthoughts. In *Themes from Kaplan*, ed. J. Almog, J. Perry & H. Wettstein, pp. 565–614. Oxford: Oxford University Press.

Kaplan, D. (1990a) Thoughts on demonstratives. In *Demonstratives*, ed. P. Yourgrau, pp. 34–49. Oxford: Oxford University Press.
Kaplan, D. (1990b) Words. *Proceedings of the Aristotelian Society Supplement*, **64**, 93–119.
Karttunen, L. & Peters, S. (1979) Conversational implicature. In *Syntax and Semantics, 11: Presupposition*, ed. C.-K. Oh & D. A. Dinneen, pp. 1–56. New York: Academic Press.
Kasher, A. (1982) Gricean inference revisited. *Philosophica*, **29**, 25–44.
Katz, J. J. (1964a) Semi-sentences. In *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 400–16. Englewood Cliffs, NJ: Prentice-Hall.
Katz, J. J. (1964b) Analyticity and contradiction in natural language. In *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 519–43. Englewood Cliffs, NJ: Prentice-Hall.
Katz, J. J. (1964c) Mentalism in linguistics. *Language*, **40**, 124–37. Reprinted in *Readings in the Philosophy of Language*, ed. J. Rosenberg & C. Travis, pp. 364–78. Englewood Cliffs, NJ: Prentice-Hall, 1971.
Katz, J. J. (1966) *Philosophy of Language*. New York: Harper and Row.
Katz, J. J. (1972) *Semantic Theory*. New York: Harper and Row.
Katz, J. J. (1974) Where things now stand with the analytic-synthetic distinction. *Synthese*, **28**, 283–319.
Katz, J. J. (1977a) A proper theory of names. *Philosophical Studies*, **31**, 1–80.
Katz, J. J. (1977b) *Propositional Structure and Illocutionary Force*. New York: Crowell.
Katz, J. J. (1986a) Why intensionalists ought not be Fregeans. In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. E. LePore, pp. 59–91. Oxford: Basil Blackwell.
Katz, J. J. (1986b) *Cogitations*. Oxford: Oxford University Press.
Katz, J. J. (1987) Descartes's *Cogito*. *Philosophical Quarterly*, **68**, 154–81. Reprinted in *Demonstratives*, ed. P. Yourgrau, pp. 154–181. Oxford: Oxford University Press, 1990.
Katz, J. J. (1990) Has the description theory been refuted? In *Meaning and Method: Essays in Honor of Hilary Putnam*, ed. G. Boolos, pp. 31–61. Cambridge: Cambridge University Press.
Katz, J. J. (1994) Names without bearers. *Philosophical Review*, **103**, 1–39.
Katz, J. J. & Fodor, J. A. (1963) The structure of a semantic theory. *Language*, **39**, 170–210. Reprinted in *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 479–518. Englewood Cliffs, NJ: Prentice-Hall, 1964.
Katz, J. J. & Langendoen, D. T. (1976) Pragmatics and presupposition. *Language*, **52**, 1–17.
Kemmerling, A. (1986) Utterer's meaning revisited. In *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, ed. R. Grandy & R. Warner, pp. 131–55. Oxford: Clarendon Press.
Kempson, R. (1975) *Presupposition and the Delimitation of Semantics*. Cambridge: Cambridge University Press.
Kiel, F. C. (1989) *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
Kim, J. (1972) Physicalism and the multiple realizability of mental states. In *Readings in Philosophy of Psychology*, ed. N. Block, vol. 1, pp. 235–6. Cambridge, MA: Harvard University Press, 1980. Originally published as Phenomenal properties, psychophysical laws, and the identity theory. *Monist*, **56**, 177–192.

Kimble, G. A. (1961) *Hilgard and Marquis' Conditioning and Learning*. New York: Appleton-Century-Crofts.
Kiteley, M. (1991) Subjectivity's bailiwick and the person of its bailiff. In *Meaning and Truth: Essential Readings in Modern Semantics*, ed. J. Garfield and M. Kiteley, pp. 372–95. New York: Paragon House.
Kittay, E. (1987) *Metaphor: Its Cognitive Force and Linguistic Structure*. Oxford: Oxford University Press.
Kneale, W. & Kneale, M. (1962) *The Development of Logic*. Oxford: Clarendon Press.
Knott, T. (1934) Standard English and incorrect English. *American Speech*, **9**, 83–9.
Koch, P. J. (1983) Expressing emotions. *Pacific Philosophical Quarterly*, **64**, 176–89.
Krech, D., Crutchfield, R. S. & Livson, N. (1969) *Elements of Psychology*, 2nd ed. New York: Knopf.
Kretzmann, N. (1967) History of semantics. In *Encyclopedia of Philosophy, Vol. 8*, ed. P. Edwards, pp. 358–406. New York: Macmillan.
Kretzmann, N. (1968) The main thesis of Locke's semantic theory. *Philosophical Review*, **77**, 175–96.
Kripke, S. (1977) Speaker reference and semantic reference. *Midwest Studies in Philosophy*, **2**, 255–78.
Kripke, S. (1979) A puzzle about belief. In *Meaning and Use*, ed. A. Margalit, pp. 239–83. Dordrecht: Reidel.
Kripke, S. (1982) *Wittgenstein on Rules and Private Languages*. Cambridge, MA: Harvard University Press.
Krucêra, H. (1982) The mathematics of language. In *The American Heritage Dictionary, Second College Edition*, ed. M. Berube, D. Neely & P. DeVinne, pp. 37–40. Boston: Houghton Mifflin.
Kuhn, T. S. (1970) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
Künne, W. (1997) First person propositions: A Fregean account. In *Direct Reference, Indexicality, and Propositional Attitudes*, ed. W. Künne, A. Newen & M. Anduschus, pp. 49–68. Stanford, CA: CSLI Publications.
Kyburg, H. (1980) Functional architecture and free will. *Behavioral and Brain Sciences*, **3**, 143–6.
Kyle, J. G. & Woll, B. (1985) *Sign Language: The Study of Deaf People and Their Language*. Cambridge: Cambridge University Press.
Labov, W. (1966) *The Social Stratification of Behavior*. Washington, DC: Center for Applied Linguistics.
Labov, W. (1970) *The Study of Nonstandard English*. Champaign, IL: National Council of Teachers of English.
Lahav, R. (1989) Against compositionality: The case of adjectives. *Philosophical Studies*, **57**, 261–79.
Lakoff, G. (1971a) On generative semantics. In *Semantics: An Interdisciplinary Reader*, ed. D. Steinberg & L. Jakobovits, pp. 232–96. Cambridge: Cambridge University Press.
Lakoff, G. (1971b) Presupposition and relative well-formedness. In *Semantics: An Interdisciplinary Reader*, ed. D. Steinberg & L. Jakobovits, pp. 329–40. Cambridge: Cambridge University Press.

Lakoff, G. (1987) *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
Lakoff, G. (1989) Some empirical results about the nature of concepts. *Mind and Language*, **4**, 103–29.
Lance, M. & O'Leary-Hawthorne, J. (1997) *The Grammar of Meaning*. Cambridge: Cambridge University Press.
Landau, S. I. (1984) *Dictionaries: The Art and Craft of Lexicography*. New York: Charles Scribner's Sons.
Large, A. (1985) *The Artificial Language Movement*. Oxford: Basil Blackwell.
Laurence, S. (1996) A Chomskian alternative to convention-based semantics. *Mind*, **105**, 269–301.
Laurence, S. & Margolis, E. (1997) Regress arguments against the language of thought. *Analysis*, **57**, 60–6.
Leech, G. (1983) *Principles of Pragmatics*. London: Longmans.
Lehrer, A. (1974) *Semantic Fields and Lexical Structure*. Amsterdam: North Holland.
Leibniz, G. W. F. (1676) What is an idea? In *Leibniz: Selections*, ed. P. P. Wiener, pp. 281–3. New York: Charles Scribner's Sons, 1951.
Leibniz, G. W. F. (1677) Dialogue on the connection between things and words. In *Leibniz: Selections*, ed. P. P. Wiener, pp. 6–11. New York: Charles Scribner's Sons, 1951.
Leibniz, G. W. F. (1684) Reflections on knowledge, truth, and ideas. In *Leibniz: Selections*, ed. P. P. Wiener, pp. 283–90. New York: Charles Scribner's Sons, 1951.
Leibniz, G. W. F. (1709) *New Essays Concerning Human Understanding*, 2nd ed., trans. A. G. Langley. Chicago: Open Court.
Leonard, S. A. (1929) *The Doctrine of Correctness in English Usage 1700–1800*. New York: Russell and Russell.
Levinson, S. C. (1983) *Pragmatics*. Cambridge: Cambridge University Press.
Lewis, D. (1969) *Convention*. Cambridge, MA: Harvard University Press.
Lewis, D. (1972) General semantics. In *Semantics of Natural Language*, ed. D. Davidson & G. Harman, pp. 169–218. Dordrecht: Reidel.
Lewis, D. (1975) Languages and language. In *Minnesota Studies in the Philosophy of Language*, ed. K. Gunderson, vol. 7, pp. 3–35. Minneapolis: University of Minnesota Press.
Lewis, D. (1976) Conventions: Reply to Jamieson. *Canadian Journal of Philosophy*, **6**, 113–20.
Lewis, D. (1979) Attitudes *de dicto* and *de se*. *Philosophical Review*, **88**, 513–43.
Lewis, D. (1981a) What puzzling Pierre does not believe. *Australasian Journal of Philosophy*, **59**, 282–9.
Lewis, D. (1981b) Index, context, and content. In *Philosophy and Grammar*, ed. S. Kanger & S. Ohman, pp. 79–100. Dordrecht: Reidel.
Lewis, D. (1986) *On the Plurality of Worlds*. Oxford: Basil Blackwell.
Lewis, D. (1992) Meaning without use: Reply to Hawthorne. *Australasian Journal of Philosophy*, **70**, 106–10.
Lewy, C. (1976) *Meaning and Modality*. Cambridge: Cambridge University Press.
Liedtke, F. (1990) Meaning and expression: Marty and Grice on intentional semantics. In *Mind, Meaning, and Metaphysics*, ed. K. Mulligan, pp. 29–49. Dordrecht: Kluwer.

Linde, C. & Labov, W. (1975) Spatial networks as a site for the study of language and thought. *Language*, **51**, 924–39.
Linsky, L. (1963) Reference and referents. In *Philosophy and Ordinary Language*, ed. C. C. Caton, pp. 74–89. Urbana: University of Illinois Press.
Linsky, L. (1972) Analytic/synthetic and semantic theory. In *Semantics of Natural Language*, ed. D. Davidson & G. Harman, pp. 473–82. Dordrecht: Reidel.
Linsky, L. (1977) *Names and Descriptions*. Chicago: University of Chicago Press.
Linsky, L. (1983) *Oblique Contexts*. Chicago: University of Chicago Press.
Lloyd, D. (1991) Leaping to conclusions: Connections, consciousness, and the computational mind. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 444–57. Dordrecht: Kluwer.
Loar, B. (1972) Reference and propositional attitudes. *Philosophical Review*, **81**, 43–62.
Loar, B. (1976a) The semantics of singular terms. *Philosophical Studies*, **30**, 353–77.
Loar, B. (1976b) Two theories of meaning. In *Truth and Meaning*, ed. G. Evans & J. McDowell, pp. 138–61. Oxford: Oxford University Press.
Loar, B. (1981) *Mind and Meaning*. Cambridge: Cambridge University Press.
Loar, B. (1987) Names in thought. *Philosophical Studies*, **51**, 169–85.
Loar, B. (1988) Social content and psychological content. In *Contents of Thought*, ed. R. H. Grimm & D. D. Merrill, pp. 99–110. Tucson: University of Arizona Press. Reprinted in *The Nature of Mind*, ed. D. M. Rosenthal, pp. 569–75. Oxford: Oxford University Press, 1991.
Loar, B. (1991) Can we explain intentionality? In *Meaning in Mind: Fodor and His Critics*, ed. B. Lower & G. Rey, pp. 119–35. Oxford: Basil Blackwell.
Locke, J. (1690) *An Essay Concerning Human Understanding*, ed. A. C. Fraser. New York: Dover, 1959.
Loewer, B. & Rey, G., eds. (1991) *Meaning in Mind: Fodor and His Critics*. Oxford: Basil Blackwell.
Lombard, L. G. & Stine, G. C. (1974) Grice's intentions. *Philosophical Studies*, **25**, 207–12.
Loux, M., ed. (1974) *Ockham's Theory of Terms: Part I of the Summa Logicae*. Notre Dame, IN: University of Notre Dame Press.
Loux, M., ed. (1979) *The Possible and the Actual*. Ithaca, NY: Cornell University Press.
Love, N. (1994) Correctness and norms. In *Encyclopedia of Language and Linguistics*, ed. R. Asher, pp. 1398–403. Oxford: Pergamon Press.
Löfstedt, E. (1959) *Late Latin*. London: Kegan Paul, Trench Trubner.
Luntley, M. (1984) The sense of a name. *Philosophical Quarterly*, **34**, 265–82.
Lycan, W. (1981) Toward a homuncular theory of believing. *Cognition and Brain Theory*, **4**, 139–59.
Lycan, W. (1984) *Logical Form in Natural Language*. Cambridge, MA: MIT Press.
Lycan, W. (1985) The paradox of naming. In *Analytical Philosophy in Comparative Perspective*, ed. B. K. Matilal & J. L. Shaw, pp. 81–102. Dordrecht: Reidel.
Lycan, W., ed. (1990) *Mind and Cognition: A Reader*. Oxford: Basil Blackwell.
Lycan, W. (1993) A deductive argument for the representational theory of thought. *Mind and Language*, **8**, 414–22.
Lycan, W. (1995) Philosophy of language. In *The Cambridge Dictionary of Philosophy*, ed. R. Audi, pp. 586–9. Cambridge: Cambridge University Press.

Lyons, J. (1971) *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
Lyons, J. (1977) *Semantics*. Cambridge: Cambridge University Press.
Macdonald, C. (1995) Introduction: Classicism v. connectionism. In *Connectionism*, ed. C. Macdonald & G. Macdonald, pp. 3–27. Oxford: Basil Blackwell.
Macdonald, C. & Macdonald, G., eds. (1995) *Connectionism: Debates in Psychological Explanation*. Oxford: Basil Blackwell.
Macià, J. (1998) On concepts and conceptions. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 175–85. Atascadero, CA: Ridgeview.
MacKay, A. F. (1972) Professor Grice's theory of meaning. *Mind*, **81**, 57–66.
Malcolm, N. (1954) Wittgenstein's *Philosophical Investigations*. *Philosophical Review*, **63**, 530–59.
Malcolm, N. (1973) Thoughtless brutes. *Proceedings and Addresses of the American Philosophical Association*, **46**, 5–20.
Maloney, J. C. (1989) *The Mundane Matter of the Mental Language*. Cambridge: Cambridge University Press.
Marconi, D. (1990) Dictionaries and proper names. *History of Philosophy Quarterly*, **7**, 77–92.
Margolis, E. (1998) How to acquire a concept. *Mind and Language*, **13**, 347–69. Reprinted in *Concepts: Core Readings*, ed. E. Margolis and S. Laurence, pp. 549–67. Cambridge, MA: MIT Press, 1999.
Margolis, E. & Laurence, S. (1999) Concepts and cognitive science. In *Concepts: Core Readings*, ed. E. Margolis & S. Laurence, pp. 3–81. Cambridge, MA: MIT Press.
Martinich, A. P. (1979) The attributive use of proper names. *Analysis*, **37**, 159–63.
Martinich, A. P. (1981) Translator's commentary. In *Thomas Hobbes: Part I of de Corpore*, ed. I. C. Hungerland & G. R. Vick, pp. 335–440. New York: Abaris.
Martinich, A. P. (1984) *Communication and Reference*. Berlin and New York: Walter de Gruyter.
Marty, A. (1908) *Untersuchungen Zur Grundlegung der Allgemeinen Grammatik und Sprach-Philosophie*. Halle: Niemeyer.
Marx, M. H., ed. (1969) *Learning: Processes*. New York: Macmillan.
Mates, B. (1950) Synonymity. *University of California Publications in Philosophy*, **25**, 201–26. Reprinted in *Semantics and the Philosophy of Language*, ed. L. Linsky, pp. 111–38. Urbana: University of Illinois Press, 1952.
Mates, B. (1958) Comments on the foregoing. *Inquiry*, **1**, 240.
Mates, B. (1986) *The Philosophy of Leibniz*. New York: Oxford University Press.
McCawley, J. D. (1968) The role of semantics in a grammar. In *Universals in Linguistic Theory*, ed. E. Bach & R. T. Harms, pp. 124–69. New York: Holt.
McCawley, J. D. (1987) The multidimensionality of pragmatics. *Behavioral and Brain Sciences*, **10**, 723–4.
McCawley, J. D. (1994) Generative semantics. In *Encyclopedia of Language and Linguistics*, ed. R. E. Asher, pp. 1398–1403. Oxford: Pergamon Press.
McClelland, J. L., Rumelhart, D. E. & Hinton, G. E. (1986) The appeal of parallel distributed processing. In *Parallel Distributed Processing*, ed. D. E. Rumelhart & J. L. McClelland, pp. 3–44. Cambridge, MA: MIT Press.
McDermott, M. (1988) The narrow semantics of names. *Mind*, **97**, 224–37.
McDowell, J. (1977) On the sense and reference of a proper name. *Mind*, **86**, 159–85.

McDowell, J. (1978) Physicalism and primitive denotation: Field on Tarski. *Erkenntnis*, **13**, 131–52. Reprinted in *Meaning and Truth: Essential Readings in Modern Semantics*, ed. J. Garfield and M. Kiteley, pp. 297–315. New York: Paragon House, 1991.

McDowell, J. (1980) Meaning, communication, and knowledge. In *Philosophical Subjects: Essays Presented to P. F. Strawson*, ed. Z. van Straaten, pp. 117–39. Oxford: Clarendon Press.

McGinn, C. (1982) The structure of content. In *Thought and Object*, ed. A. Woodfield, pp. 207–58. Oxford: Clarendon Press.

McGinn, C. (1997) *The Character of Mind*. Oxford: Oxford University Press.

McKay, T. (1981) On proper names in belief ascriptions. *Philosophical Studies*, **39**, 287–303.

McKinsey, M. (1978) Names and intentionality. *Philosophical Review*, **87**, 171–200.

McKinsey, M. (1994) Individuating beliefs. In *Philosophical Perspectives, 8: Logic and Language*, ed. J. Tomberlin, pp. 303–29. Atascadero, CA: Ridgeview.

Meggle, G. (1997) *Brundbegriffe de Kommunicakation*, 2nd ed. Berlin: Walter de Gruyter.

Meinong, A. (1904) The theory of objects. In *Realism and the Background of Phenomenology*, ed. R. M. Chisholm, pp. 76–117. Glencoe, IL: The Free Press, 1960.

Meinong, A. (1910) *On Assumptions*, ed. and trans. J. Heanue. Berkeley: University of California Press, 1983.

Mesthrie, R. (1994) Language maintenance, shift, and death. In *Encyclopedia of Language and Linguistics*, ed. R. E. Asher, pp. 1988–93. Oxford: Pergamon Press.

Mey, J. (1993) *Pragmatics: An Introduction*. Oxford: Basil Blackwell.

Mill, J. (1829) *Analysis of the Phenomenon of the Human Mind*, ed. J. S. Mill. London: Longmans Green Reader and Dyer, 1869.

Mill, J. S. (1843) *A System of Logic*. London: Longmans, Green.

Mill, J. S. (1865) *An Examination of the Philosophy of Sir William Hamilton*. London: Longmans.

Mill, J. S. (1869) Notes. In *James Mill: Analysis of the Phenomenon of the Human Mind*, ed. J. S. Mill, vol. 1, London: Longmans Green Reader and Dyer.

Mill, J. S. (1879) *A System of Logic*, 10th ed. London: Longmans, Green.

Miller, S. (1986) Conventions, interdependence of action, and collective ends. *Noûs*, **20**, 117–40.

Miller, S. (1992) On conventions. *Australasian Journal of Philosophy*, **70**, 435–44.

Millikan, R. (1984) *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Millikan, R. (1986) Thoughts without laws: Cognitive science without content. *Philosophical Review*, **95**, 47–80.

Millikan, R. (1987) What Peter thinks when he hears Mary speak. *Behavioral and Brain Sciences*, **10**, 725–6.

Millikan, R. (1998a) A more plausible kind of "recognitional concept." In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 35–41. Atascadero, CA: Ridgeview.

Millikan, R. (1998b) A common structure for concepts of individuals, stuffs, and real kinds: More mama, more milk, and more mouse. *Behavioral and Brain Sciences*, **21**, 55–65. Reprinted in *Concepts: Core Readings*, ed. E. Margolis & S. Laurence, pp. 525–47. Cambridge, MA: MIT Press, 1999.

Minsky, M. (1975) A framework for representing knowledge. In *The Psychology of Computer Vision*, ed. P. H. Winston. New York: McGraw Hill.
Moore, G. E. (1942) A reply to my critics. In *The Philosophy of G. E. Moore*, ed. P. A. Schilpp, pp. 533–689. La Salle, IL: Open Court, 1968.
Moore, J. G. (1999a) Misdisquotation and substitutivity: When not to infer belief from assent. *Mind*, **108**, 335–65.
Moore, J. G. (1999b) Saving substitutivity in simple sentences. *Analysis*, **59**, 91–105.
Morris, C. (1946) *Signs, Language, and Behaviour*. Englewood Cliffs, NJ: Prentice-Hall.
Moya, C. (1998) Boghossian's *reductio* of compatibilism. In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 243–51. Atascadero, CA: Ridgeview.
Muller, S. H. (1964) *The World's Living Languages*. New York: Ungar.
Murphy, G. & Medin, D. (1985) The role of theories in conceptual coherence. *Psychological Review*, **92**, 289–316. Reprinted in *Concepts: Core Readings*, ed. E. Margolis and S. Laurence, pp. 425–58. Cambridge, MA: MIT Press, 1999.
Nagel, T. (1974) What is it like to be a bat? *The Philosophical Review*, **83**, 435–50.
Neale, S. (1992) Paul Grice and the philosophy of language. *Linguistics and Philosophy*, **15**, 509–59.
Nelson, R. J. (1992) *Naming and Reference*. London: Routledge and Kegan Paul.
Newman, J. (1962) Review of *Webster's Third New International Dictionary*. *The Quarterly Journal of Speech*, **68**, 437–8.
Nisbett, R. & Ross, L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.
Noonan, H. (1980–1) Names and belief. *Proceedings of the Aristotelian Society*, **81**, 93–108.
Noonan, H. (1984) Fregean thoughts. *Philosophical Quarterly*, **34**, 205–24.
Nunberg, G. (1981) Validating pragmatic explanations. In *Radical Pragmatics*, ed. P. Cole, pp. 199–222. New York: Academic Press.
Nunberg, G. (1982) English and good English. In *The American Heritage Dictionary*, ed. M. Berube, D. Neely & P. DeVinne, pp. 34–6. Boston: Houghton Mifflin.
Nurse, D. & Spear, T. (1985) *The Swahili*. Philadelphia: University of Pennsylvania Press.
Ockham, W. *Expositio super duos libros perihermenias*, trans. P. Boehner. In *Philosophical Writings: A Selection*, ed. P. Boehner. Indianapolis, IN: Bobbs-Merrill, 1964.
Ockham, W. *Ordinatio*, trans. P. Boehner. In *Philosophical Writings: A Selection*, ed. P. Boehner. Indianapolis, IN: Bobbs-Merrill, 1964.
Ockham, W. *Quodlibeta*, trans. P. Boehner. In *Philosophical Writings: A Selection*, ed. P. Boehner. Indianapolis, IN: Bobbs-Merrill, 1964.
Ockham, W. (1325) *Summa Logicae, Part I: On Terms*. In *Ockham's Theory of Terms: Part I of the Summa Logicae*, ed. M. J. Loux, pp. 49–221. Notre Dame, IN: University of Notre Dame Press, 1974.
Osherson, D. & Smith, E. (1981) On the adequacy of prototype theory as a theory of concepts. *Cognition*, **9**, 35–58. Reprinted in *Concepts: Core Readings*, ed. E. Margolis & S. Laurence, pp. 261–78. Cambridge, MA: MIT Press, 1999.
O'Grady, W., Dobrovolsky, M. & Aronoff, M. (1993) *Contemporary Linguistics: An Introduction,* 2nd ed. New York: St. Martin's Press.

O'Leary-Hawthorne, J. (1989) On the compatibility of connectionist and classical models. *Philosophical Psychology*, **2**, 5–15.
O'Leary-Hawthorne, J. (1990) A note on "languages and language." *Australasian Journal of Philosophy*, **68**, 116–18.
O'Leary-Hawthorne, J. (1993) Meaning and evidence: Reply to Lewis. *Australasian Journal of Philosophy*, **71**, 202–12.
Pap, A. (1955) Belief, synonymity, and analysis. *Philosophical Studies*, **6**, 11–15.
Pap, A. (1957) Belief and propositions. *Philosophy of Science*, **24**, 123–36.
Pappas, G. S. (1995) Idea. In *The Cambridge Dictionary of Philosophy*, ed. R. Audi, p. 355. Cambridge: Cambridge University Press.
Partee, B. (1984) Compositionality. In *Varieties of Formal Semantics*, ed. F. Landman & F. Veltman, pp. 281–311. Dordrecht: Foris.
Pateman, T. (1982) David Lewis's theory of convention and the social life of language. *Journal of Pragmatics*, **6**, 135–57.
Patton, T. E. (1994) Proper names: Possibility and culture. In *Language, Mind, and Art: Essays in Appreciation of Analysis, and in Honor of Paul Ziff*, ed. D. Jamieson, pp. 39–53. Dordrecht: Kluwer.
Patton, T. E. & Stampe, D. W. (1969) The rudiments of meaning: On Ziff on Grice. *Foundations of Language*, **5**, 2–16.
Pavlov, I. P. (1927) *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. London: Oxford University Press.
Pawley, A. (1985) Lexicalization. In *Language and Linguistics: The Interdependence of Theory, Data, and Application*, ed. D. Tannen & J. E. Alatis, pp. 98–120. Washington, DC: Georgetown University Press.
Peacocke, C. (1975) Proper names, reference, and rigid designation. In *Meaning, Reference, and Necessity*, ed. S. Blackburn, pp. 109–32. Cambridge: Cambridge University Press. Reprinted in *Definite Descriptions: A Reader*, ed. G. Osterling, pp. 201–34. Cambridge MA: MIT Press, 1998.
Peacocke, C. (1976) Truth definitions and actual languages. In *Truth and Meaning: Essays in Semantics*, ed. G. Evans & J. McDowell, pp. 162–88. Oxford: Clarendon Press.
Peacocke, C. (1981) Demonstrative thought and psychological explanation. *Synthese*, **49**, 187–217.
Peacocke, C. (1983) *Sense and Content: Experience, Thought, and Their Relations*. Oxford: Oxford University Press.
Peacocke, C. (1986) *Thoughts: An Essay on Content*. Oxford: Basil Blackwell.
Peacocke, C. (1989) Possession conditions: A focal point for theories of concepts. *Mind and Language*, **4**, 51–6.
Peacocke, C. (1992) *A Study of Concepts*. Cambridge, MA: MIT Press.
Pelczar, M. & Rainsbury, J. (1998) The indexical character of names. *Synthese*, **114**, 293–317.
Perry, D. L. (1967) *The Concept of Pleasure*. The Hague: Mouton.
Perry, J. (1977) Frege on demonstratives. *Philosophical Review*, **86**, 474–97.
Perry, J. (1979) The problem of the essential indexical. *Noûs*, **13**, 3–21. Reprinted in *Meaning and Truth: The Essential Readings in Modern Semantics*, ed. J. Garfield and M. Kiteley, pp. 613–27. New York: Paragon House, 1991.

Perry, J. (1980) A problem about continued belief. *Pacific Philosophical Quarterly*, **61**, 317–32.
Perry, J. (1983) Castañeda on He and I. In *Agent, Language, and the Structure of the World*, ed. J. Tomberlin, pp. 15–39. Atascadero, CA: Ridgeview.
Perry, J. (1990) Self-notions. *Logos*, **11**, 17–31.
Perry, J. (1997a) Indexicals and demonstratives. In *A Companion to the Philosophy of Language*, ed. B. Hale & C. Wright, pp. 586–612. Oxford: Oxford University Press.
Perry, J. (1997b) Reflexivity, indexicality, and names. In *Direct Reference, Indexicality, and Propositional Attitudes*, ed. W. Künne, A. Newen, and M. Anduschus, pp. 3–19. Stanford, CA: CSLI Publications.
Peters, R. (1956) *Hobbes*. Harmondsworth: Penguin.
Peters, R. (1967) *Hobbes*, 2nd ed. Harmondsworth: Penguin.
Pettit, P. (1987) Inference and information. *Behavioral and Brain Sciences*, **10**, 727–9.
Peirce, C. S. (1931–5) *Collected Papers*, ed. P. Hartshorne, P. Weiss & A. W. Burks. Cambridge, MA: Harvard University Press.
Pitcher, G. (1964) Introduction. In his *Truth*, pp. 1–15. Englewood Cliffs, NJ: Prentice-Hall.
Place, U. T. (1956) Is consciousness a brain process? *British Journal of Psychology*, **47**, 44–50.
Plantinga, A. (1985) Self-profile and replies. In *Alvin Plantinga*, ed. J. E. Tomberlin & P. Van Inwagen, pp. 3–97, 313–94. Dordrecht: Reidel.
Plato. *Cratylus*, trans. B. Jowett. In *The Collected Dialogues of Plato*, ed. E. Hamilton, pp. 421–74. Princeton, NJ: Princeton University Press, 1961.
Plato. *Phaedo*, trans. H. Tredennick. In *The Collected Dialogues of Plato*, ed. E. Hamilton, pp. 40–98. Princeton, NJ: Princeton University Press, 1961.
Plato. *Theaetetus*, trans. F. M. Cornford. In *The Collected Dialogues of Plato*, ed. E. Hamilton, pp. 845–919. Princeton, NJ: Princeton University Press, 1961.
Platts, M. (1979) *Ways of Meaning*. London: Routledge and Kegan Paul.
Pollock, J. L. (1980) Thinking about an object. *Midwest Studies in Philosophy*, **5**, 487–99.
Posner, R. (1980) Semantics and pragmatics of sentence connectives in natural languages. In *Speech-Act Theory and Pragmatics*, ed. J. Searle, F. Kiefer & M. Bierwisch, pp. 169–203. Amsterdam: Reidel.
Posner, R. (1997) Romance languages. In *Encyclopaedia Britannica*, pp. 620–39. Chicago: Encyclopaedia Britannica.
Postal, P. M. (1964) Limitations of phrase structure grammars. In *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 137–51. Englewood Cliffs, NJ: Prentice-Hall.
Price, H. H. (1953) *Thinking and Experience*. Cambridge, MA: Harvard University Press.
Price, H. H. (1954) Belief and will. *Proceedings of the Aristotelian Society Supplement*, **28**, 1–26.
Price, H. H. (1969) *Belief*. New York: Humanities.
Priestley, J. (1762) *A Course of Lectures on the Theory of Language and Universal Grammar*. Menston, England: Scolar Press, 1970.
Putnam, H. (1954) Synonymity, and the analysis of belief sentences. *Analysis*, **14**, 114–22.

Putnam, H. (1967) The nature of mental states. In *The Nature of Mind*, ed. D. M. Rosenthal, pp. 197–210. Oxford: Oxford University Press, 1991. Originally published as "Psychological Predicates" in *Art, Mind, and Religion*, ed. W. H. Capitan & D. D. Merrill, pp. 37–8. Pittsburgh: University of Pittsburgh Press.

Putnam, H. (1970a) On properties. In *Essays in Honor of Carl G. Hempel*, ed. A. R. Anderson, P. Benacerraf, A. Grünbaum, G. Massey, N. Rescher, and R. Rudner. Dordrecht: Reidel. Reprinted in his *Philosophical Papers*, pp. 305–22. Cambridge: Cambridge University Press, 1975.

Putnam, H. (1970b) Is semantics possible? In *Languages, Belief, and Metaphysics*, ed. H. Kiefer & M. Munitz. New York: State University of New York Press. Reprinted in *Concepts: Core Readings*, ed. E. Margolis & S. Laurence, pp. 178–87. Cambridge, MA: MIT Press, 1999.

Putnam, H. (1973) Meaning and reference. *The Journal of Philosophy*, **70**, 699–711. Reprinted in *The Philosophy of Language*, ed. A. P. Martinich, pp. 308–15. Oxford: Oxford University Press, 1991.

Putnam, H. (1975) The meaning of "meaning." In *Language, Mind, and Knowledge*, ed. K. Gunderson, pp. 131–93. Minneapolis: University of Minnesota Press.

Putnam, H. (1981) *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Pylyshyn, Z. (1980) Computation and cognition: Issues in the foundations of cognitive science. *The Behavioral and Brain Sciences*, **3**, 111–32, 154–69.

Quillian, M. R. (1968) Semantic memory. In *Semantic Information Processing*, ed. M. L. Minsky, pp. 216–55. Cambridge, MA: MIT Press.

Quine, W. V. O. (1936) Truth by convention. In *Essays for A. N. Whitehead*, ed. O. H. Lee. New York: Longmans. Reprinted in *Ways of Paradox*, ed. W. V. O. Quine, pp. 70–9. New York: Random House, 1966.

Quine, W. V. O. (1948) On what there is. In his *From a Logical Point of View*, pp. 1–19. New York: Harper and Row, 1963.

Quine, W. V. O. (1953) Two dogmas of empiricism. In his *From a Logical Point of View*, pp. 20–46. New York: Harper and Row, 1963. Reprinted in *Readings in the Philosophy of Language*, ed. J. F. Rosenberg and C. Travis, pp. 63–81. Englewood Cliffs, NJ: Prentice-Hall, 1971.

Quine, W. V. O. (1956) Quantifiers and propositional attitudes. *Journal of Philosophy*, **53**. Reprinted in *The Ways of Paradox*, ed. W. V. O. Quine, pp. 183–94. New York: Random House, 1966.

Quine, W. V. O. (1959) *Methods of Logic*. New York: Hold, Rinehart, and Winston.

Quine, W. V. O. (1960) *Word and Object*. Cambridge, MA: MIT Press.

Quine, W. V. O. (1961) The problem of meaning in linguistics. In his *From a Logical Point of View*, 2nd ed., pp. 47–64. Cambridge, MA: Harvard University Press, 1961. Reprinted in *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 21–32. Englewood Cliffs, NJ: Prentice-Hall, 1964.

Quine, W. V. O. (1967) Russell's ontological development. In *Bertrand Russell, Philosopher of the Century*, ed. R. Schoenman. Boston: Little Brown.

Quine, W. V. O. (1969a) Ontological relativity. In *Ontological Relativity and Other Essays*, pp. 26–68. New York: Columbia University Press.

Quine, W. V. O. (1969b) Foreword to *Convention* by D. K. Lewis, pp. xi–xii. Cambridge, MA: Harvard University Press.
Quine, W. V. O. (1972) Methodological reflections on current linguistic theory. In *Semantics of Natural Language*, ed. D. Davidson & G. Harman, pp. 442–54. Dordrecht: Reidel.
Quine, W. V. O. (1987) Ideas. In *Quiddities: An Intermittently Philosophical Dictionary*, pp. 87–9. Cambridge, MA: Harvard University Press.
Quintilian, M. F. *Institutio Oratoria*, trans. H. E. Butler. In *Landmarks in Linguistic Thought I*, ed. R. Harris & T. J. Taylor, pp. 60–75. London: Routledge, 1997.
Radutzky, E. (1993) The education of deaf people in Italy and the use of Italian Sign Language. In *Deaf History Unveiled*, ed. J. V. Van Cleve, pp. 237–51. Washington, DC: Gallaudet University Press.
Ramsey, W., Stich, S. P. & Garon, J. (1991) Connectionism, eliminativism, and the future of folk psychology. In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich & D. E. Rumelhart, pp. 199–228. Hillsdale, NJ: Erlbaum.
Rand, A. (1969) *Introduction to Objectivist Epistemology*. New York: The Objectivist.
Read, A. W. (1962) Review of *Webster's Third New International Dictionary*. *The Quarterly Journal of Speech*, **68**, 438–9.
Recanati, F. (1986) On defining communicative intentions. *Mind and Language*, **1**, 213–42.
Recanati, F. (1987) *Meaning and Force: The Pragmatics of Performative Utterances*. Cambridge: Cambridge University Press.
Recanati, F. (1989) Referential/attributive: A contextualist proposal. *Philosophical Studies*, **56**, 217–49.
Recanati, F. (1990) Direct reference, meaning, and thought. *Noûs*, **24**, 697–722.
Recanati, F. (1993) *Direct Reference: From Language to Thought*. Oxford: Basil Blackwell.
Recanati, F. (1995) The alleged priority of literal interpretation. *Cognitive Science*, **19**, 207–32.
Reid, T. (1764) *Inquiry into the Human Mind*. In *Thomas Reid's Inquiry and Essays*, ed. K. Lehrer & E. Beanblossom, pp. 3–125. Indianapolis, IN: Bobbs-Merrill, 1975.
Reid, T. (1785) *Essays on the Intellectual Powers of Man*, ed. B. Brody. Cambridge, MA: MIT Press, 1969.
Reimer, M. (1995) A defense of *de re* belief reports. *Mind and Language*, **10**, 446–63.
Reimer, M. (1998) The Wettstein/Salmon debate: Critique and resolution. *Pacific Philosophical Quarterly*, **79**, 130–51.
Rey, G. (1991) An explanatory budget for connectionism and eliminativism. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 219–40. Dordrecht: Kluwer.
Rey, G. (1994) Concepts. In *A Companion to the Philosophy of Mind*, ed. S. Guttenplan. Oxford: Basil Blackwell.
Rey, G. (1995) A not "merely empirical" argument for a language of thought. *Philosophical Perspectives*, **9**, 201–22.
Richard, M. (1981) Temporalism and eternalism. *Philosophical Studies*, **39**, 1–13.
Richard, M. (1987) Attitude ascriptions, semantic theory, and pragmatic evidence. *Proceedings of the Aristotelian Society*, **87**, 243–62.

Richard, M. (1989) How I say what you think. *Midwest Studies in Philosophy*, **14**, 317–37.
Richard, M. (1990) *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. Cambridge: Cambridge University Press.
Rickford, J. (1992) Pidgins and creoles. In *International Encyclopedia of Linguistics, Vol. 2*, ed. W. Bright, pp. 224–9. Oxford: Oxford University Press.
Rieber, S. (1997) Conventional implicatures as tacit performatives. *Linguistics and Philosophy*, **20**, 51–72.
Rips, L. (1995) The current status of research on concept combination. *Mind and Language*, **10**, 72–104.
Robins, R. H., Maurer, D. W., & Ivic, P. (1997) Language. In *Encyclopedia Britannica*, pp. 548–71. Chicago: Encyclopedia Britannica.
Rorty, R. (1970) Incorrigibility as the mark of the mental. *Journal of Philosophy*, **67**, 399–424.
Rosch, E. & Mervis, C. B. (1975) Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573–605.
Rosenberg, J. (1974) *Linguistic Representation*. Dordrecht: Reidel.
Rosenberg, J. (1993) Another look at proper names. In *Philosophical Perspectives, 7: Language and Logic*, ed. J. Tomberlin, pp. 505–30. Atascadero, CA: Ridgeview.
Ruhl, C. (1989) *On Monosemy: A Study in Linguistic Semantics*. Albany: State University of New York Press.
Rumelhart, D. E. & McClelland, J. L. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press.
Russell, B. (1905) On denoting. In *The Philosophy of Language*, 2nd ed., ed. A. P. Martinich, pp. 203–11. Oxford: Oxford University Press, 1990. From *Mind*, **14**, 479–93.
Russell, B. (1910–11) Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society*, **11**, 108–28. Reprinted in his *Mysticism and Logic*, pp. 209–32. London: Allen and Unwin, 1917.
Russell, B. (1921) *The Analysis of Mind*. London: Allen and Unwin.
Russell, B. (1940) *An Inquiry into Meaning and Truth*. London: Allen and Unwin.
Russell, S. (1987) Rationality as an explanation of language. *Behavioral and Brain Sciences*, **10**, 730–1.
Ryle, G. (1949) *The Concept of Mind*. New York: Barnes and Noble.
Ryle, G. (1951) Thinking and language. From *Proceedings of the Aristotelian Society Supplement,* **25**, 65–82. Reprinted in *Collected Essays*, vol. 2, pp. 258–71. New York: Barnes and Noble, 1971.
Ryle, G. (1953) Thinking. *Acta Psychologica*, **9**, 294–300. Reprinted in *Collected Essays*, vol. 2, pp. 294–300. New York: Barnes and Noble, 1971.
Ryle, G. (1955) Categories. In *Logic and Language*, ed. A. Flew, pp. 75–81. Oxford: Basil Blackwell.
Ryle, G. (1957) The theory of meaning. In *British Philosophy in the Mid-Century*, ed. C. A. Mace, pp. 239–64. London: Allen and Unwin.
Ryle, G. (1958) A puzzling element in the notion of thinking. From *Proceedings of the British Academy*, **44**, 129–44. Reprinted in *Collected Essays*, vol. 2, pp. 391–406. New York: Barnes and Noble, 1971.

Ryle, G. (1962) Thinking thoughts and having concepts. From *Logique et Analyse*, **20**, 446–50. Reprinted in *Collected Essays*, vol. 2, pp. 446–50. New York: Barnes and Noble, 1971.
Ryle, G. (1979) *On Thinking*. Totowa, NJ: Rowman and Littlefield.
Sachs, J. S. (1967) Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, **2**, 437–42.
Sadock, J. M. (1974) *Toward a Linguistic Theory of Speech Acts*. New York: Academic Press.
Salmon, N. (1981) *Reference and Essence*. Princeton, NJ: Princeton University Press.
Salmon, N. (1982) Assertion and incomplete definite descriptions. *Philosophical Studies*, **42**, 37–45.
Salmon, N. (1986) *Frege's Puzzle*. Cambridge, MA: MIT Press.
Salmon, N. (1989a) Tense and singular propositions. In *Themes from Kaplan*, ed. J. Almog, J. Perry & H. Wettstein, pp. 331–92. Oxford: Oxford University Press.
Salmon, N. (1989b) How to become a Millian heir. *Noûs*, **23**, 211–20.
Salmon, N. (1989c) Reference and information content: Names and descriptions. In *Handbook of Philosophical Logic, Volume IV*, ed. D. Gabbay & F. Guenthner, pp. 409–61. Dordrecht: Reidel.
Salmon, N. (1990) A Millian heir rejects the wages of *sinn*. In *Propositional Attitudes*, ed. C. A. Anderson & J. Owens, pp. 215–57. Stanford, CA: CSLI Publications.
Samet, J. (1986) Troubles with Fodor's nativism. In *Midwest Studies in Philosophy*, *Volume X: Studies in the Philosophy of Mind*, ed. P. A. French, T. E. Uehling & H. K. Wettstein, pp. 575–94. Minneapolis: University of Minnesota Press.
Samet, J. & Flanagan, O. (1989) Innate representation. In *Rerepresentation: Readings in the Philosophy of Mental Representation*, ed. S. Silver, pp. 189–210. Dordrecht: Kluwer.
Saul, J. (2001) Review of *Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory* by Wayne Davis. *Noûs*, **35**, 630–41.
Saussure, F. (1916) *Course in General Linguistics*, trans. W. Baskin. New York: McGraw-Hill, 1966.
Schank, R. & Abelson, R. (1977) *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Erlbaum.
Scheffler, I. (1955) On synonymy and indirect discourse. *Philosophy of Science*, **22**, 39–44.
Schein, J. D. & Stewart, D. A. (1996) *Language in Motion: Exploring the Nature of Sign*. Washington, DC: Gallaudet University Press.
Schelling, T. (1960) *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Schiffer, S. (1972) *Meaning*. Oxford: Clarendon Press.
Schiffer, S. (1978) The basis of reference. *Erkenntnis*, **13**, 171–206.
Schiffer, S. (1981) Indexicals and the theory of reference. *Synthese*, **49**, 43–100.
Schiffer, S. (1982) Intention-based semantics. *Notre Dame Journal of Formal Logic*, **23**, 119–56.
Schiffer, S. (1987a) *Remnants of Meaning*. Cambridge, MA: MIT Press.
Schiffer, S. (1987b) The 'Fido'-Fido theory of belief. In *Philosophical Perspectives, 1: Metaphysics*, ed. J. Tomberlin, pp. 455–80. Atascadero, CA: Ridgeview.

Schiffer, S. (1990) The mode-of-presentation problem. In *Propositional Attitudes*, ed. C. A. Anderson & J. Owens, pp. 249–68. Stanford, CA: CSLI Publications.
Schiffer, S. (1991) Does mentalese have a compositional semantics? In *Meaning in Mind: Fodor and His Critics*, ed. B. Loewer & G. Rey, pp. 181–99. Oxford: Basil Blackwell.
Schiffer, S. (1992) Belief ascription. *Journal of Philosophy*, **89**, 499–521.
Schiffer, S. (1994) A paradox of meaning. *Noûs*, **28**, 279–324.
Schiffer, S. (1995) Descriptions, indexicals, and belief reports: Some dilemmas (but not the ones you expect). *Mind*, **104**, 107–31.
Schiffrin, D. (1994) *Approaches to Discourse*. Oxford: Basil Blackwell.
Schwarz, D. (1979) *Naming and Referring: The Semantics and Pragmatics of Singular Terms*. New York: De Gruyter.
Searle, J. (1969) *Speech-Acts*. Cambridge: Cambridge University Press.
Searle, J. (1979) *Expression and Meaning*. Cambridge: Cambridge University Press.
Searle, J. (1983) *Intentionality*. Cambridge: Cambridge University Press.
Searle, J. (1986) Meaning, communication, and representation. In *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, ed. R. Grandy & R. Warner, pp. 209–26. Oxford: Clarendon Press.
Searle, J. (1997) The explanation of cognition. In *Thought and Language*, ed. J. Preston, pp. 103–26. Cambridge: Cambridge University Press.
Sellars, W. (1955) Putnam on synonymity and belief. *Analysis*, **15**, 117–20.
Sellars, W. (1958) The Chisholm-Sellars correspondence on intentionality. In *Minnesota Studies in the Philosophy of Science*, vol. 2, ed. M. Feigl, M. Scriven & G. Maxwell, pp. 529–39. Minneapolis: University of Minnesota Press. Reprinted in *Intentionality, Mind, and Language*, ed. A. Marras, pp. 214–48. Urbana: University of Illinois Press, 1972.
Sellars, W. (1963) Empiricism and the philosophy of mind. In *Science, Perception, and Reality*, pp. 127–96. London: Routledge and Kegan Paul.
Sellars, W. (1969) Language as thought and as communication. *Philosophy and Phenomenological Research*, **29**, 506–27. Reprinted in *Essays in Philosophy and History*, pp. 93–117. Dordrecht: Reidel, 1974.
Sellars, W. (1979) *Naturalism and Ontology*. Reseda, CA: Ridgeview.
Shoemaker, S. (1984) *Identity, Cause, and Mind*. Cambridge: Cambridge University Press.
Silverberg, A. (1992) Putnam on functionalism. *Philosophical Studies*, **67**, 111–31.
Simpson, J. M. Y. (1994) Language. In *Encyclopedia of Language and Linguistics*, ed. R. E. Asher, pp. 1893–7. Oxford: Pergamon Press.
Siple, P., ed. (1978) *Understanding Language through Sign Language Research*. New York: Academic Press.
Skinner, B. F. (1938) *The Organization of Behavior*. New York: Appleton-Century-Crofts.
Skinner, B. F. (1957) *Verbal Behavior*. New York: Appleton-Century-Crofts.
Skulsky, H. (1986) Metaphorese. *Noûs*, **20**, 351–69. Reprinted in *Meaning and Truth*, ed. J. Garfield & M. Kiteley, pp. 582–98. New York: Paragon House, 1991.
Smart, J. J. C. (1959) Sensations and brain processes. *Philosophical Review*, **58**, 141–56.

Smith, D. V. et al. (1952) *The English Language Arts*. New York: Appleton-Century-Crofts. Prepared by the Commission on the English Curriculum of the National Council of Teachers of English.
Smith, D. W. (1981) Indexical sense and reference. *Synthese*, **49**, 101–28.
Smith, D. W. (1982) Husserl on demonstrative reference and perception. In *Husserl, Intentionality, and Cognitive Science: Recent Studies in Phenomenology*, ed. H. Dreyfus, pp. 193–214. Cambridge, MA: MIT Press.
Smith, E. E. (1988) Concepts and thought. In *The Psychology of Human Thought*, ed. R. J. Sternberg & E. E. Smith, pp. 19–49. Cambridge: Cambridge University Press.
Smith, E. E. & Medin, D. L. (1981) *Categories and Concepts*. Cambridge, MA: Harvard University Press.
Smith, N. V. & Wilson, D. (1979) *Modern Linguistics: The Results of Chomsky's Revolution*. Harmondsworth: Penguin.
Smith, P. T. (1994) Thought and language. In *Encyclopedia of Language and Linguistics*, ed. R. E. Asher, pp. 4610–15. Oxford: Pergamon Press.
Smolensky, P. (1988) On the proper treatment of connectionsm. In *Connectionism*, ed. C. Macdonald & G. Macdonald, pp. 28–89. Oxford: Basil Blackwell. From *Behavioral and Brain Sciences*, **11**, 1–74.
Soames, S. (1987a) Substitutivity. In *On Being and Saying: Essays for Richard Cartwright*, ed. J. J. Thompson, pp. 99–132. Cambridge, MA: MIT Press.
Soames, S. (1987b) Direct reference, propositional attitudes, and semantic content. *Philosophical Topics*, **15**, 197–239.
Sober, E. (1998) Innate knowledge. In *Routledge Encyclopedia of Philosophy*, ed. E. J. Craig, pp. 794–7. London: Routledge.
Sosa, D. (1996) The import of a puzzle about belief. *Philosophical Review*, **105**, 373–402.
Sosa, E. (1970) Propositional attitudes *de dicto* and *de re*. *Journal of Philosophy*, **67**, 883–96.
Sosa, E. (1983) Propositions and indexical attitudes. In *On Believing: Epistemological and Semiotic Approaches*, ed. H. Parret, pp. 316–31. New York: Walter de Gruyter.
Sosa, E. (1995) *De re* belief, action explanations, and the essential indexical. In *Modality, Morality, and Belief*, ed. W. Sinnott-Armstrong, D. Raffman & N. Asher, pp. 235–49. Cambridge: Cambridge University Press.
Sperber, D. & Wilson, D. (1986a) *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.
Sperber, D. & Wilson, D. (1986b) Loose talk. *Proceedings of the Aristotelian Society*, **86**, 153–71. Reprinted in *Pragmatics: A Reader*, ed. S. Davis, pp. 540–9. Oxford: Oxford University Press, 1991.
Sperber, D. & Wilson, D. (1987) Précis of *Relevance: Communication and Cognition*. *Behavioral and Brain Sciences*, **10**, 679–754.
Stalnaker, R. (1976a) Possible worlds. *Noûs*, **10**, 65–75. Reprinted in *The Possible and the Actual*, ed. M. J. Loux, pp. 225–34. Ithaca, NY: Cornell University Press, 1979.
Stalnaker, R. (1976b) Propositions. In *Issues in the Philosophy of Language*, ed. A. MacKay & D. Merrill, pp. 79–91. New Haven, CT: Yale University Press. Reprinted in *Meaning and Truth*, ed. J. Garfield and M. Kiteley, pp. 467–77. New York: Paragon, 1991.

Stalnaker, R. (1998) What might nonconceptual content be? In *Philosophical Issues, 9: Concepts*, ed. E. Villanueva, pp. 339–52. Atascadera, CA: Ridgeview.
Stampe, D. W. (1968) Towards a grammar of meaning. *Philosophical Review*, **77**, 137–74.
Stampe, D. W. (1979) Towards a causal theory of linguistic representation. In *Contemporary Perspectives in the Philosophy of Language*, ed. P. French, T. Uehling & H. Wettstein, pp. 81–102. Minneapolis: University of Minnesota Press.
Stanley, J. (1997) Names and rigid designation. In *A Companion to the Philosophy of Language*, ed. B. Hale & C. Wright, pp. 555–85. Oxford: Oxford University Press.
Sterelny, K. (1989) Fodor's nativism. *Philosophical Studies*, **55**, 119–41.
Sterelny, K. (1990) *The Representational Theory of Mind*. Oxford: Basil Blackwell.
Stich, S. (1975) Introduction: The idea of innateness. In his *Innate Ideas*, pp. 1–22. Berkeley: University of California Press.
Stich, S. (1980) Computation without representation. *Behavioral and Brain Sciences*, **3**, 152.
Stich, S. (1983) *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
Stich, S. (1986) Are belief predicates systematically ambiguous? In *Belief: Form, Content, and Function*, ed. R. Bogdan, pp. 119–47. Oxford: Clarendon Press.
Stillings, N. A., Feinstein, M. H., Garfield, J. L., Rissland, E. L., Rosenbaum, D. A., Weisler, S. E. & Baker-Ward, L. (1987) *Cognitive Science: An Introduction*. Cambridge, MA: MIT Press.
Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L. & Rissland, E. L. (1995) *Cognitive Science: An Introduction*, 2nd ed. Cambridge, MA: MIT Press.
Stocker, M. (1987) Emotional thoughts. *American Philosophical Quarterly*, **24**, 59–69.
Stout, G. F. (1899) *A Manual of Psychology*. London: W. B. Clive.
Strawson, P. F. (1964) Intention and convention in speech acts. *Philosophical Review*, **73**, 439–60. Reprinted in *Pragmatics: A Reader*, ed. S. Davis, pp. 290–302. Oxford: Oxford University Press, 1991.
Strawson, P. F. (1971) Meaning and truth. In *Logico-Linguistic Papers*, pp. 170–89. London: Methuen.
Suppes, P. (1986) The primacy of utterer's meaning. In *Philosophical Grounds of Rationality: Intentions, Categories, Ends.*, ed. R. Grandy & R. Warner, pp. 109–29. Oxford: Clarendon Press.
Taschek, W. (1995) On belief content and that-clauses. *Mind and Language*, **10**, 274–98.
Taylor, C. (1980) Review of *Linguistic Behavior*, by J. Bennett. *Dialogue*, **19**, 290–301.
Thomason, R. (1990) Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In *Intentions in Communication*, ed. P. R. Cohen, J. L. Morgan & M. Pollack, pp. 325–63. Cambridge, MA: MIT Press.
Thorndike, E. L. (1911) *Animal Intelligence*. New York: Macmillan.
Tienson, J. (1991) Introduction. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 1–29. Dordrecht: Kluwer.
Titchener, E. B. (1897) *Outline of Psychology*, 2nd ed. New York: Macmillan.
Titchener, E. B. (1909) *Lectures on the Experimental Psychology of the Thought Processes*. New York: Macmillan.
Titchener, E. B. (1914) *A Primer of Psychology*. New York: Macmillan.

Titchener, E. B. (1929) *Systematic Psychology: Prolegomena*. Ithaca, NY: Cornell University Press.
Tomberlin, J. (1991) Belief, self-ascription, and ontology. In *Philosophical Issues, 1: Consciousness*, ed. E. Villanueva, pp. 233–59. Atascadero, CA: Ridgeview.
Tormey, A. (1971) *The Concept of Expression*. Princeton, NJ: Princeton University Press.
Travis, C. (1997) Pragmatics. In *A Companion to the Philosophy of Language*, ed. B. Hale & C. Wright, pp. 87–107. Oxford: Basil Blackwell.
Tsohatzidis, S. L. (1994) Ways of doing things with words. In *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, ed. S. L. Tsohatzidis, pp. 1–25. London: Routledge.
Tversky, A. & Kahneman, D. (1973) Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207–32.
Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. In *Judgment under Uncertainty: Heuristics and Biases*, ed. D. Kahneman, P. Slovic & A. Tversky, pp. 3–20. Cambridge: Cambridge University Press.
Ullmann-Margalit, E. (1977) *The Emergence of Norms*. Oxford: Oxford University Press.
Urmson, J. O. (1967) Ideas. In *Encyclopedia of Philosophy*, ed. P. Edwards, vol. 4, pp. 118–21. New York: Macmillan.
Urmson, J. O. (1968) Criteria of intensionality. *Proceedings of the Aristotelian Society Supplement*, **42**, 107–22.
van Gelder, T. (1990) Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, **14**, 355–84.
van Gelder, T. (1991a) Classical questions, radical answers: Connectionism and the structure of mental representations. In *Connectionism and the Philosophy of Mind*, ed. T. Horgan & J. Tienson, pp. 355–81. Dordrecht: Kluwer.
van Gelder, T. (1991b) What is the "D" in "PDP"? A survey of the concept of distribution. In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich & D. E. Rumelhart, pp. 33–59. Hillsdale, NJ: Erlbaum.
van Heijenoort, J. (1967) Gödel's theorem. In *Encyclopedia of Philosophy*, ed. P. Edwards, vol. 3, pp. 348–57. New York: Macmillan.
Vanderveken, D. (1990) *Meaning and Speech Acts*. Cambridge: Cambridge University Press.
Vendler, Z. (1972) *Res Cogitans: An Essay in Rational Psychology*. Ithaca, NY: Cornell University Press.
Vendler, Z. (1976) Thinking of individuals. *Noûs*, **10**, 35–46.
Vendler, Z. (1977) Words in thought. *Philosophical Exchange*, **2**, 55–66.
Vlach, F. (1981) Speaker's meaning. *Linguistics and Philosophy*, **4**, 359–92.
Vogelin, C. & Vogelin, F. (1988) Languages of the world. In *Encyclopaedia Britannica*, ed. R. McHenry, pp. 591–815. Chicago: Encyclopaedia Britannica.
Walker, R. (1975) Conversational implicatures. In *Meaning, Reference, and Necessity: New Studies in Semantics*, ed. S. Blackburn, pp. 133–81. Cambridge: Cambridge University Press.
Wallace, J. (1977) Only in the context of a sentence do words have any meaning. *Midwest Studies in Philosophy*, **2**, 305–25.

Warnock, G. J. (1953) *Berkeley*. Harmondsworth: Pelican.
Warren, H. C. (1921) *A History of the Association Psychology*. New York: Scribners.
Watson, J. B. (1924) *Psychology from the Standpoint of a Behaviorist*, 2nd ed. Philadelphia: Lippincott.
Watson, J. B. (1930) *Behaviorism*. Chicago: University of Chicago Press.
Weiskrantz, L., ed. (1988) *Thought without Language*. Oxford: Clarendon Press.
Weiskrantz, L. (1997) Thought without language: Thought without awareness? In *Thought and Language*, ed. J. Preston, pp. 127–50. Cambridge: Cambridge University Press.
Wells, J. C. (1994) Esperanto. In *The Encyclopedia of Language and Linguistics*, ed. R. Asher, pp. 1143–5. Oxford: Pergamon Press.
Wettstein, H. K. (1976) Can what is asserted be a sentence? *Philosophical Review*, **85**, 196–207. Reprinted in *Has Semantics Rested on a Mistake? and Other Essays*, ed. H. Wettstein, pp. 9–19. Palo Alto, CA: Stanford University Press, 1991.
Wettstein, H. K. (1981) Demonstrative reference and definite descriptions. *Philosophical Studies*, **40**, 241–57. Reprinted in *Has Semantics Rested on a Mistake? and Other Essays*, ed. H. Wettstein, pp. 35–58. Palo Alto, CA: Stanford University Press, 1991.
Wettstein, H. K. (1986) Has semantics rested on a mistake? *Journal of Philosophy*, **83**, 185–209.
Wettstein, H. K. (1991) *Has Semantics Rested on a Mistake? and Other Essays*. Palo Alto, CA: Stanford University Press.
Wetzel, L. (1989) That numbers could be objects. *Philosophical Studies*, **56**, 273–92.
Wetzel, L. (2002) On types and words. *Journal of Philosophical Research*, **27**, 239–65.
Wetzel, L. (forthcoming) *Of Types and Tokens*. Cambridge, MA: MIT Press.
Whorf, B. (1956) *Language, Thought, and Reality*. Cambridge, MA: MIT Press.
Wierzbicka, A. (1992a) Semantic primitives and semantic fields. In *Frames, Fields, and Contrasts*, ed. A. Lehrer & E. Kittay, pp. 209–27. Hillsdale, NJ: Erlbaum.
Wierzbicka, A. (1992b) The semantics of interjection. *Journal of Pragmatics*, **18**, 159–92.
Wiggins, D. (1971) On sentence-sense, word-sense, and difference of word-sense. Towards a philosophical theory of dictionaries. In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, ed. D. Steinberg & L. Jakobovits, pp. 14–34. Cambridge: Cambridge University Press.
Wilbur, R. B. (1987) *American Sign Language: Linguistic and Applied Dimensions*. Boston: Little, Brown.
Williams, C. J. F. (1991) You and she. *Analysis*, **51**, 143–6.
Wilson, D. & Sperber, D. (1981) On Grice's theory of conversation. In *Conversation and Discourse*, ed. P. Werth, pp. 155–78. New York: St. Martin's Press.
Wilson, N. L. (1970) Grice on meaning: The ultimate counterexample. *Noûs*, **4**, 295–302.
Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul, 1972.
Wittgenstein, L. (1953) *Philosophical Investigations*. Oxford: Oxford University Press.
Woodfield, A. (1982) On specifying the contents of thought. In *Thought and Object*, ed. A. Woodfield, pp. 259–97. Oxford: Clarendon Press.

Woodfield, A. (1991) Conceptions. *Mind*, **100**, 547–72.
Woodfield, A. (1997) Social externalism and conceptual diversity. In *Thought and Language*, ed. J. Preston, pp. 77–102. Cambridge: Cambridge University Press.
Woodward, J. (1978) Historical bases of American Sign Language. In *Understanding Language through Sign Language Research*, ed. P. Siple, pp. 333–47. New York: Academic Press.
Woodworth, R. S. (1921) *Psychology: A Study of Mental Life*. New York: Holt.
Woodworth, R. S. & Sells, S. B. (1935) An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology*, **18**, 451–60.
Wreen, M. (1998) Proper names and the necessity of identity statements. *Synthese*, **114**, 319–35.
Wright, R. A. (1975) Meaning$_{nn}$ and conversational implicature. In *Syntax and Semantics, 3: Speech Acts*, ed. P. Cole & J. L. Morgan, pp. 363–82. New York: Academic Press.
Wundt, W. (1911) *An Introduction to Psychology*, trans. R. Pintner. London: George Allen, 1912.
Yagisawa, T. (1984) The pseudo-Mates argument. *Philosophical Review*, **93**, 407–18.
Yagisawa, T. (1993) A semantic solution to Frege's puzzle. In *Philosophical Perspectives, 7: Language and Logic*, ed. J. Tomberlin, pp. 135–54. Atascadero, CA: Ridgeview.
Yagisawa, T. (1998) Naming and its place in reference. *Lingua e Stile*, **33**, 445–58.
Yoder, C. H., Suydam, F. H. & Snavely, F. A. (1975) *Chemistry*. New York: Harcourt Brace Jovanovich.
Yolton, J. W. (1956) *John Locke and the Way of Ideas*. Oxford: Oxford University Press.
Yourgrau, P., ed. (1990) *Demonstratives*. Oxford: Oxford University Press.
Yu, P. (1979) On the Gricean program about meaning. *Linguistics and Philosophy*, **3**, 273–388.
Zadrozny, W. (1994) From compositional to systematic semantics. *Linguistics and Philosophy*, **17**, 329–42.
Zalta, E. N. (1989) Singular propositions, abstract constituents, and propositional attitudes. In *Themes from Kaplan*, ed. J. Almog, J. Perry & H. Wettstein, pp. 455–78. Oxford: Oxford University Press.
Ziff, P. (1960) *Semantic Analysis*. Ithaca, NY: Cornell University Press.
Ziff, P. (1964) On understanding "understanding utterances." In *The Structure of Language*, ed. J. Fodor & J. J. Katz, pp. 390–9. Englewood Cliffs, NJ: Prentice-Hall.
Ziff, P. (1967) On H. P. Grice's account of meaning. *Analysis*, **28**, 1–8.

Index