"Concepts and Cognitive Science" in Concepts: Core Readings (MIT Press, 1999), pp. 3-81.

Chapter 1

Concepts and Cognitive Science

Stephen Laurence and Eric Margolis

1. *Introduction: Some Preliminaries*

Concepts are the most fundamental constructs in theories of the mind. Given their importance to all aspects of cognition, it's no surprise that concepts raise so many controversies in philosophy and cognitive science. These range from the relatively local

> Should concepts be thought of as bundles of features, or do they embody mental theories?

to the most global

> Are concepts mental representations, or might they be abstract entities?

Indeed, it's even controversial whether concepts are objects, as opposed to cognitive or behavioral abilities of some sort. Because of the scope of the issues at stake, it's inevitable that some disputes arise from radically different views of what a theory of concepts ought to achieve—differences that can be especially pronounced across disciplinary boundaries. Yet in spite of these differences, there has been a significant amount of interdisciplinary interaction among theorists working on concepts. In this respect, the theory of concepts is one of the great success stories of cognitive science. Psychologists and linguists have borrowed freely from philosophers in developing detailed empirical theories of concepts, drawing inspiration from Wittgenstein's discussions of family resemblance, Frege's distinction between sense and reference, and Kripke's and Putnam's discussions of externalism and essentialism. And philosophers have found psychologists' work on categorization to have powerful implications for a wide range of philosophical debates. The philosopher Stephen Stich (1993) has gone so far as to remark that current empirical models in psychology undermine a traditional approach to philosophy in which philosophers engage in conceptual analyses. As a consequence of this work, Stich and others have come to believe that philosophers have to rethink their approach to topics in areas as diverse as the philosophy of mind and ethics. So even if disciplinary boundaries have generated the appearance of disjoint research, it's hard to deny that significant interaction has taken place.

We hope this volume will underscore some of these achievements and open the way for increased cooperation. In this introduction, we sketch the recent history of theories of concepts. However, our purpose isn't solely one of exposition. We also provide a number of reinterpretations of what have come to be standard arguments in the field and develop a framework that lends more prominence to neglected areas

This paper was fully collaborative; the order of the authors' names is arbitrary.

of the intellectual geography. Given the vast range of theories at play, it would be impossible to say anything substantive without offending some theoretical scruples. So we should say right now that we don't claim to be completely neutral. As we go along, we try to justify our choices to some extent, but inevitably, in a space as short as this, certain views will receive less attention. Our strategy is to present what we take to be the main theories of concepts and do this in terms of idealized characterizations that provide rather rough yet useful demarcations.

Before we begin, however, there are three preliminary issues that need to be mentioned. Two can be dealt with fairly quickly, but the third—concerning the ontological status of concepts—requires a more extended treatment.

Primitive, Complex and Lexical Concepts[1]

For a variety of reasons, most discussions of concepts have centered around *lexical concepts*. Lexical concepts are concepts like BACHELOR, BIRD, and BITE—roughly, ones that correspond to lexical items in natural languages.[2] One reason for the interest in lexical concepts is that it's common to think that words in natural languages inherit their meanings from the concepts they are used to express. In some discussions, concepts are taken to be just those mental representations that are expressed by words in natural languages. However, this usage is awkward, since it prohibits labeling as concepts those representations that are expressed by complex natural language expressions. One wouldn't be able to say, for example, that the concept BLACK CAT (corresponding to the English expression "black cat") is composed of the simpler concepts BLACK and CAT; only the latter would be concepts. Yet most of the reasons that one would have to single out BLACK and CAT and the like as concepts apply equally to complexes that have these as their constituents. There may be little difference between lexical concepts and other complex concepts apart from the fact that the former are lexicalized; indeed, on many views, lexical concepts are themselves complex representations. At the same time, it seems wrong to designate as concepts mental representations of any size whatsoever. Representations at the level of complete thoughts—that is, ones that may express whole propositions—are too big to be concepts. Accordingly, we will take *concepts* to be subpropositional mental representations.

Two other points of terminology should be mentioned. We'll say that *primitive concepts* are ones that lack structure. *Complex concepts*, in contrast, are concepts that aren't primitive. In the cognitive science literature, primitive concepts are sometimes called *atomic* concepts or *features*, although this terminology is confused by the fact that "feature" is sometimes used more permissively (i.e., to refer to any component of a concept) and is sometimes used more restrictively (i.e., to refer to only primitive sensory concepts). We'll adopt a permissive use of "feature" and say that unstruc-

1. Throughout, we will refer to concepts by using expressions in small caps. When quoting, we will adjust other people's notations to our own.

2. For present purposes, there is no need to insist on a more precise characterization, apart from noting that the concepts in question are ones that are usually encoded by single morphemes. In particular, we won't worry about the possibility that one language may use a phrase where another uses a word, and we won't worry about exactly what a word is (but for some alternative conceptions, see Di Sciullo and Williams 1987). Admittedly, the notion of a lexical concept isn't all that sharp, but it does help to orient the discussion toward the specific concepts that have been most actively subjected to investigation, for instance, BIRD as opposed to BIRDS THAT EAT REDDISH WORMS IN THE EARLY MORNING HOURS.

tured concepts are primitive or atomic. What exactly it means to say that a concept has, or lacks, structure is another matter. This brings us to our second preliminary point.

Two Models of Conceptual Structure

Most theories of concepts treat lexical concepts as structured complexes. This raises the issue of what it is for such representational complexes to have structure. Despite the important role that conceptual structure plays in many debates, there has been little explicit discussion of this question. We discern two importantly different models of structure that are implicit in these debates.

The first view we'll call the *Containment Model.* On this view, one concept is a structured complex of other concepts just in case it literally has those other concepts as proper parts. In this way, a concept *C* might be composed of the concepts *X*, *Y*, and *Z*. Then an occurrence of *C* would necessarily involve an occurrence of *X*, *Y*, and *Z*; because *X*, *Y*, and *Z* are contained within *C*, *C* couldn't be tokened without *X*, *Y*, and *Z* being tokened. For example, the concept DROPPED THE ACCORDION couldn't be tokened without ACCORDION being tokened. As an analogy, you might think of the relation that words bear to phrases and sentences. The word "accordion" is a structural element of the sentence "Tony dropped the accordion" in the sense that it is a proper part of the sentence. Consequently, you can't utter a token of the sentence "Tony dropped the accordion" without thereby uttering a token of the word "accordion."

The second view, which we'll can the *Inferential Model,* is rather different. According to this view, one concept is a structured complex of other concepts just in case it stands in a privileged relation to these other concepts, generally, by way of some type of inferential disposition. On this model, even though *X*, *Y*, and *Z* may be part of the structure of *C*, *C* can still occur without necessitating their occurrence. For example, RED might have a structure implicating the concept COLOR, but on the Inferential Model, one could entertain the concept RED without having to token the concept COLOR. At most, one would have to have certain dispositions linking RED and COLOR—for example, the disposition to infer *X* IS COLORED from *X* IS RED.

Thus, for any claim that a concept has such-and-such structure—or such-and-such *type* of structure (see sec. 7)—there will be, in principle, two possible interpretations of the claim: one in terms of the Containment Model and one in terms of the Inferential Model. The significance of these distinctions will become clearer once we present some specific theories of concepts. For now we simply want to note that discussions of conceptual structure are often based on an implicit commitment to one of these models and that a proper evaluation of a theory of concepts may turn on which model is adopted.

Concepts as Abstracta vs. Concepts as Mental Representations

The third and last preliminary point that we need to discuss concerns a more basic issue—the ontological status of concepts. In accordance with virtually all discussions of concepts in psychology, we will assume that concepts are mental particulars. For example, your concept GRANDMOTHER is a mental representation of a certain type, perhaps a structured mental representation in one of the two senses we've isolated. It should be said, however, that not all theorists accept as their starting point the thesis that concepts are mental particulars. In philosophy especially it's not uncommon to

think of concepts as abstract entities.[3] Clarifying the motivations for this view and its relation to standard psychological accounts requires a digression.[4] We hope the reader will bear with us, however, since some of the distinctions that are at play in this dispute will be relevant later on.

Perhaps the best way to begin is by way of the nineteenth-century German philosopher Gottlob Frege and his distinction between *sense* and *reference*. Frege was primarily interested in language, in particular, artificial languages used in logic, mathematics, and science. But the distinctions he drew have analogues for natural language and theories about the nature of mental representation.

In the first instance, it helps to think of senses in terms of another technical notion in Frege—the *mode of presentation* for the referent of a term. Frege discussed a variety of cases where different terms refer to the same object but do so by characterizing the object in different ways. For instance, "two plus two" and "the square root of 16" both refer to the number four, but they incorporate different ways of characterizing it. This distinction—between referent and mode of presentation—is standardly applied to expressions of every size and semantic category. We can speak of the mode of presentation for a name, or a kind term, or even a whole sentence, just as we can for a phrase. "Mark Twain" and "Samuel Clemens" may refer to the same individual, but their modes of presentation for this individual aren't the same. Similarly, "gold" and "element with atomic number 79" may refer to the same stuff, but clearly under distinct modes of presentation.

The connection with senses is that Frege held that expressions have a sense, in addition to a referent, and that the sense of an expression "contains" the mode of presentation for its referent. We needn't worry about all of the details here, but to get clearer about senses, it pays to think of them as being characterized by the roles that Frege asked them to play. Three ought to be clearly distinguished (cf. Burge 1977):

1. *Senses are the cognitive content of linguistic expressions* This role is related to what has come to be known as *Frege's Puzzle*. Frege asks how two identity statements—"the morning star is the morning star" and "the morning star is the evening star"—could differ in cognitive content. Both are identity statements involving coreferential terms denoting the planet Venus, yet the first is a truism, the second a significant astronomical discovery. Frege's solution to the puzzle is to say that the expressions involved in these statements have senses, and the differences in cognitive content correspond to differences between the senses they express.
2. *Senses determine reference* For Frege, our linguistic and conceptual access to the world is mediated by the senses of the expressions in our language. A sense, as a mode of presentation, fixes or determines the referent of an expression. And it is through our grasp of a sense that we access the referent. The

3. Yet another alternative is the view that concepts are not particulars at all but are, instead, behavioral or psychological abilities. We take it that behavioral abilities are ruled out for the same reasons that argue against behaviorism in general (see, e.g., Chomsky 1959). However, the view that concepts are psychological abilities is harder to evaluate. The chief difficulty is that more needs to be said about the nature of these abilities. Without a developed theory, it's not even clear that an appeal to abilities is in conflict with the view that concepts are particulars. For example, such abilities might require that one be in possession of a mental particular that is deployed in a characteristic way.

4. A variety of theoretical perspectives treat concepts as abstracta, but we take the version we discuss to be representative.

expression "the morning star" refers to the object it does because this expression has the sense it does.

3. *Senses are the indirect referents of expressions in intensional contexts* Certain linguistic contexts (e.g., "... believes that ..." and other propositional attitude reports) have distinctive and peculiar semantic properties. Outside of these contexts, one can freely substitute coreferential terms without affecting the truth value of the sentence ("the morning star is bright" → "the evening star is bright"), but within these contexts, the same substitutions are not possible ("Sue believes that the morning star is bright" ↛ "Sue believes that the evening star is bright"). Frege's explanation of this type of case is that in such contexts expressions do not refer to their customary referents, but rather to their customary *senses*. Since the expressions have different customary senses, they actually have different referents in these contexts. Thus Frege is able to maintain the principle that coreferring terms can be substituted one for the other without a change in truth value, despite what otherwise may have appeared to be a decisive counterexample to the principle.

Frege's semantic theory, and the phenomena he used to motivate it, have generated a great deal of controversy, and they have had an enormous influence on the development of semantic theories in philosophy and linguistics. For now, though, the important issue is the ontological status of senses. Frege argued that senses, construed in terms of these theoretical roles, cannot be mental entities. Since it's common in philosophy to hold that concepts just are Fregean senses, it would seem that Frege's case against mental entities is especially pertinent. The problem, in his view, is that mental entities are subjective, whereas senses are supposed to be objective. Two people "are not prevented from grasping the same sense; but they cannot have the same idea" (1892/1966, p. 60). (Note that for Frege, ideas are mental entities.)

If this is the argument against the view that concepts are mental representations, however, it isn't the least bit convincing. To see why, one has to be careful about teasing apart several distinctions that can get lumped together as a single contrast between the subjective and the objective. One of these concerns the difference between mental representations, thoughts, and experiences, on the one hand, and extra-mental entities on the other. In this sense, a stone is objective, but a mental representation of a stone is subjective; it's subjective simply because it's mental. Notice, however, that subjectivity of this kind doesn't preclude the sharing of a mental representation, since two people can have the same type of mental representation. What isn't possible is for two people to have the very same *token* representation. This brings us to a second subjective-objective distinction. It can be put this way: Mental representations are subjective in that their tokens are uniquely possessed; they belong to one and only one subject. Their being subjective in this sense, however, doesn't preclude their being shareable in the relevant sense, since, again, two people can have the same representation by each having tokens of the same type. When someone says that two people have the same concept, there is no need to suppose that she is saying that they both possess the same token concept. It would make as much sense to say that two people cannot utter the same sentence because they cannot both produce the same token sentence. Clearly what matters for being able to utter the same sentence, or entertain the same concept, is being able to have tokens of the same type. So while mental representations are subjective in the two senses

we've isolated, this doesn't stop them from being objective in the sense of being shareable.[5]

In short, we see no reason why concepts can't be mental representations. And given the role of mental representations in theories of psychological processing, it would be entirely natural to follow psychological usage in calling these representations concepts. Still, this usage isn't meant to preclude a role for the abstracta that Fregeans mean to highlight. To see this, one need only consider the question of whether Frege himself could have it both ways, employing mental representations and senses. The answer, of course, is that he could. On this model, beliefs and other propositional attitudes would involve token mental representations that have other representations—concepts—as their constituents. Senses would come in as the semantic values of these representations. That is, in addition to having worldly objects and properties as their referents, mental representations (like words, on Frege's original account) would have senses too. In this way, senses help to type mental representation; they provide part of the conditions for individuating concepts.

Given this way of combining the more traditional philosophical account of concepts with the representationalism of psychology, it's little more than a terminological debate whether representations or the abstracta should be called concepts. Since we think there needn't be any confusion on this point—and since we are primarily interested in the mental representations—we'll continue to follow standard psychological usage, according to which concepts are representations.[6]

With these preliminaries out of the way, we can now turn to the theories of concepts themselves. We will work though five that figure prominently in discussions in linguistics, philosophy, and psychology. They differ in their motivations and the problems they face, but they aren't nearly as distinct from one another as is often assumed. We'll see, for example, that some problems aren't tied to a single theory; rather they present a general challenge to nearly any theory of concepts. Similarly, some of the resources that trace back to one account of concepts can be enlisted in surprising ways to help other accounts. In general, the theories that we will discuss differ in what they say about the structure of concepts. Along the way, we'll mention a number of respects in which the options regarding conceptual structure can be expanded. In the concluding section (sec. 7), we'll bring some of these strands together by discussing four ways of construing what theories of concepts have to say about the nature of concepts.

2. *The Classical Theory of Concepts*

2.1. *Concepts and Definitions*

In one way or another, most theories of concepts can be seen as reactions to, or developments of, what is known as the *Classical Theory of Concepts*.[7] The Classical

5. A third sense in which mental entities may be subjective—also suggested by Frege's text—is that they are highly idiosyncratic. Much of Frege's criticism of "ideas" is that they are too variable from one person to the next. "A painter, a horseman, and a zoologist will probably connect different ideas with the name 'Bucephalus'" (59). At best, however, Frege's observation establishes only that ideas aren't likely to be shared, not that they are, in principle, unshareable. Moreover, it's hard to see how the idiosyncrasy of ideas would motivate the claim that concepts are abstracta.

6. For further discussion on this point, see the appendix (sec. 8) and Margolis and Laurence (ms).

7. Also called the *Traditional Theory* or the *Definition View*.

Theory holds that most concepts—especially lexical concepts—have definitional structure. What this means is that most concepts encode necessary and sufficient conditions for their own application.[8] Consider, for example, the concept BACHELOR. According to the Classical Theory, we can think of this concept as a complex mental representation that specifies necessary and sufficient conditions for something to be a bachelor. So BACHELOR might be composed of a set of representations such as IS NOT MARRIED, IS MALE, and IS AN ADULT. Each of these components specifies a condition that something must meet in order to be a bachelor, and anything that satisfies them all thereby counts as a bachelor. These components, or features, yield a semantic interpretation for the complex representation in accordance with the principles of a compositional semantics.

This conception of concepts has a long history in philosophy. The seventeenth-century philosopher John Locke seems to be assuming a version of the Classical Theory when he gives his account of the concepts SUN and GOLD (1690/1975, pp. 298–299 and p. 317, respectively):

> [T]he *Idea* of the *Sun*, What is it, but an aggregate of those several simple *Ideas*, Bright, Hot, Roundish, having a constant regular motion, at a certain distance from us, and, perhaps, some other....
>
> [T]he greatest part of the *Ideas*, that make our complex *Idea* of *Gold*, are Yellowness, great Weight, Ductility, Fusibility, and Solubility, in *Aqua Regia, etc.* all united together in an unknown *Substratum*...[9]

On the Classical Theory, most concepts—including most lexical concepts—are complex representations that are composed of structurally simpler representations. What's more, it's natural to construe their structure in accordance with the Containment Model, where the components of a complex concept are among its proper parts.[10] Some of these components may themselves be complex, as in the case of BACHELOR. But eventually one reaches a level of primitive representations, which are undefined. Traditionally, these primitive representations have been taken to be sensory or perceptual in character, along broadly empiricist lines.

It is, of course, an oversimplification to speak of *the* Classical Theory of concepts, as though there were just a single, unitary theory to which all classical theorists subscribe. In reality, there is a diverse family of theories centered around the idea that

8. By "application" we mean a semantic relation; that is, a concept encodes the conditions that are singly necessary and jointly sufficient for something to be in its extension. Another sense of the term is to indicate a psychological process in which an object is judged to fall under a concept. We'll try to avoid this ambiguity by always using "application" in the semantic sense, unless the context makes it very clear that the psychological sense is intended. Notice, then, that in the first instance we have characterized the Classical Theory in semantic terms. This doesn't mean, however, that the theory is devoid of psychological import. See the discussion of concept acquisition and categorization, below.

9. Locke's views about natural kind concepts are complicated by the fact that he took natural kinds to have both a nominal and a real essence. For Locke, the real essence of a kind like gold isn't known, but the nominal essence is, and must be, in order to possess the corresponding concept. Arguably, however, he takes the nominal essence to give necessary and sufficient conditions for the application of a kind concept, since he holds that the nominal essence is defined relative to the real essence in such a way that the two track one another.

10. It's natural, but not mandatory. Alternatively, one could think of a classically structured concept as a node that stands in inferential relations to its defining features. The advantage of the Containment Model is that it makes especially clear which associated concepts are its defining features and which are incidental.

concepts have definitional structure. What we call the Classical Theory of concepts is an idealized account that abstracts away from many of their differences. To mention just one point on which classical theorists disagree: Many recent classical theorists have abandoned the strict empiricist view that concepts are ultimately composed of features expressing sensory properties.

It would be difficult to overstate the historical predominance of the Classical Theory. Aspects of the theory date back to antiquity (see Plato 1981 [chapter 2 in this volume]).[11] And the first serious challenges to its status weren't until the 1950s in philosophy, and the 1970s in psychology. Why has the Classical Theory been held in such high regard? The theory has powerful explanatory resources, offering unified accounts of concept acquisition, categorization, epistemic justification, analytic entailment, and reference determination, all of which flow directly from its basic commitments (see Fodor, J. A. et al. 1980 [chapter 21]). We will briefly review these accounts, since it helps to flesh out the Classical Theory and its substantial motivations.

Box 1

The Classical Theory

Most concepts (esp. lexical concepts) are structured mental representations that encode a set of necessary and sufficient conditions for their application, if possible, in sensory or perceptual terms.

Concept Acquisition If a concept is a complex representation built out of features that encode necessary and sufficient conditions for its application, then the natural model of concept acquisition is one where the learner acquires a concept by assembling its features. If, in accordance with the empiricist version of the Classical Theory, we add the further stipulation that primitive features are sensory or perceptual, the model we arrive at is something like the following. Through perception, sensory properties are monitored so that their representations are joined in a way that reflects environmental contingencies. Having noticed the way these properties correlate in her environment, the learner assembles a complex concept that incorporates the relevant features in such a way that something falls under the new, complex concept just in case it satisfies those features. In this way, all concepts in the end would be defined in terms of a relatively small stock of sensory concepts. As John Locke put it in *An Essay Concerning Human Understanding* (1690/1975, p. 166),

> [E]ven *the most abstruse* Ideas, how remote soever they may seem from Sense, or from any operation of our own Minds, are yet only such, as the Understanding frames to it self, by repeating and joining together *Ideas*, that it had either from Objects of Sense, or from its own operations about them....

A somewhat more recent advocate of this position is the influential twentieth-century German philosopher Rudolf Carnap. In "The Elimination of Metaphysics through Logical Analysis of Language," Carnap writes (1932/1959, pp. 62–63),

11. When, for the first time, we refer to a chapter that is reprinted in the present volume, we'll indicate this with brackets. Subsequent references will omit the bracketed material.

> In the case of many words, specifically in the case of the overwhelming majority of scientific words, it is possible to specify their meaning by reduction to other words ("constitution," definition). E.g., "'arthropodes' are animals with segmented bodies and jointed legs." ... In this way every word of the language is reduced to other words and finally to the words which occur in the so-called "observation sentences" or "protocol sentences."[12]

In the face of repeated failures to analyze everyday concepts in terms of a purely sensory base, contemporary theorists have often relaxed the strong empiricist assumption that all simple concepts must be sensory. For example, Eve Clark (1973) sees the process of acquiring the meaning of a word like "brother" as comprising several stages where semantic components get added to an initial representation. In the earliest stage the representation consists of only two components: +MALE, −ADULT. In subsequent stages, −ADULT is changed to ±ADULT, +SIBLING is added, and +RECIPROCAL is added. In this way, a representation for "brother" is gradually constructed from its constituent representations, which collectively provide a definition of the word and distinguish it from related words, such as "boy." Though these components may not be primitive, Clark isn't committed to the idea that further decomposition will always lead to purely sensory concepts. In fact, she says that many words, especially relational terms, require possibly irreducible features that encode "functional, social, or cultural factors" (p. 106). Similarly, the linguist and philosopher Jerrold Katz writes (1972 [chapter 4 in this volume], p. 40),

> [T]he English noun "chair" can be decomposed into a set of concepts which might be represented by the semantic markers in (4.10):
>
> (4.10) OBJECT, PHYSICAL, NON-LIVING, ARTIFACT, FURNITURE, PORTABLE, SOMETHING WITH LEGS, SOMETHING WITH A BACK, SOMETHING WITH A SEAT, SEAT FOR ONE.

He adds that these semantic markers—or features—require further analysis, but, like Clark, he isn't committed to a reduction that yields a purely sensory base.

No doubt, a component-by-component model of concept acquisition is compelling even when it is detached from its empiricist roots. The simplicity and power of the model provides considerable motivation for pursuing the Classical Theory.

Categorization The Classical Theory offers an equally compelling model of categorization (i.e., the application of a concept, in the psychological sense; see note 8). In fact, the model of categorization is just the ontogeny run backwards; that is, something is judged to fall under a concept just in case it is judged to fall under the features that compose the concept. So, something might be categorized as falling under the concept CHAIR by noting that it has a seat, back, legs, and so on. Categorization on this model is basically a process of checking to see if the features that are part of a concept are satisfied by the item being categorized. As with the general model of concept acquisition, this model of categorization is powerful and intuitively appealing, and it's a natural extension of the Classical Theory.

12. Throughout we'll ignore certain differences between language and thought, allowing claims about words to stand in for claims about concepts. Carnap's account is about the semantics of linguistic items but otherwise is a useful and explicit version of the Classical Theory.

Epistemic Justification A number of philosophical advocates of the Classical Theory have also emphasized the role it could play as a theory of epistemic justification. The idea is that one would be justified in taking an item to fall under a given concept by determining whether its defining components are satisfied.

The quotation from Carnap (above) is part of a larger passage where he explains that we are justified in taking a thing, *x*, to be an arthropode if a sentence of the form "the thing *x* is an arthropode" is "deducible from premises of the form '*x* is an animal,' '*x* has a segmented body,' '*x* has jointed legs' ..." (1932/1959, p. 63). Since the components that enter into the concept provide a definition of the concept, verifying that these components are satisfied is tantamount to verifying that the defined concept is satisfied as well. And since it's often assumed that the ultimate constituents of each concept express sensory properties, the verification procedure for a concept's primitive features is supposed to be unproblematic. The result is that justification for abstract or complicated concepts—including the "theoretical" concepts of science—reduces to a series of steps that implicate procedures with little epistemic risk.

Analyticity and Analytic Inferences Another important motivation for the Classical Theory is its ability to explain a variety of semantic phenomena, especially analytic inferences. Intuitively, there is a significant difference between the inferences in (1) and (2):

(1) Smith is an unmarried man. So Smith is a man.

(2) Smith is a weight-lifter. So Smith is a man.

In (1), unlike (2), the conclusion that Smith is a man seems to be guaranteed by the premise. Moreover, this guarantee seems to trace back to the meaning of the key phrase in (1), namely, "unmarried man."

Traditionally, analytic inferences have been taken to be inferences that are based on meaning, and a sentence or statement has been taken to be analytic just in case its truth is necessitated by the meanings of its constituent terms. Much of this conception of analyticity is captured in Immanuel Kant's account of analyticity as conceptual containment. "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" (1787/1965, p. 48). One of the most widely cited examples in the contemporary literature is the concept BACHELOR. Consider (3):

(3) Smith is a bachelor. So Smith is a man.

The inference in (3) is not only correct but seems to be guaranteed by the fact that it is part of the meaning of "bachelor" that bachelors are men. It's not as if one has to do a sociological study. The Classical Theory explains why one needn't look to the world in assessing (3), by claiming that the concept BACHELOR has definitional structure that implicates the concepts MAN, UNMARRIED, and so on. Thus (3) and (1) turn out to be similar, under analysis.

Katz (1972) gives much the same explanation of the validity of the inferences from (4.13)

(4.13) There is a chair in the room.

to (4.14)–(4.21)

(4.14) There is a physical object in the room.

(4.15) There is something nonliving in the room.

(4.16) There is an artifact in the room.

(4.17) There is a piece of furniture in the room.

(4.18) There is something portable in the room.

(4.19) There is something having legs in the room.

(4.20) There is something with a back in the room.

(4.21) There is a seat for one person in the room.

According to Katz, all of these inferences are to be explained by reference to the concept CHAIR and its definition, given above as (4.10). The definition is supposed to be understood in Kantian terms, by supposing that the one concept—CHAIR—contains within it the other concepts that secure the inferences—ARTIFACT, PHYSICAL OBJECT, and so on. The only difference, then, between (1) and (3), or (1) and the inferences from (4.13) to (4.14–4.21), is that the logical form of (1) is manifest, whereas the forms underlying the other inferences are hidden.[13]

Reference Determination One of the most important properties of concepts is that they are semantically evaluable. A thought may be true or false, depending on how things are with that portion of the world which the thought is about. In like fashion, an item may fall under a concept or not, depending on the concept's referential properties. When someone categorizes something as a bird, for example, she may or may not be right. This is perhaps the most basic feature of what is called the *normativity of meaning*. Just because she applies the concept BIRD to the item (in the sense that she judges it to be a bird) doesn't mean that the concept truly applies to the item (in the sense that the item is in the extension of the concept BIRD).

The referential properties of a concept are among its most essential properties. When one acquires the concept ROBIN, doing so crucially involves acquiring a concept that *refers* to robins. And when one draws an inference from ROBIN to IS A BIRD, or IS AN ANIMAL, one draws an inference *about* robins. This isn't to say that reference is sufficient to distinguish between concepts. TRIANGULAR and TRILATERAL refer to exactly the same class of mathematical objects, yet they are different concepts for all that. And in Plato's time, one might have believed that PIETY and ACTING IN A WAY THAT IS PLEASING TO THE GODS are coextensive—perhaps even necessarily coextensive—but that doesn't make them the same concept. Thus Plato can sensibly ask whether an action is pious because it is pleasing to the gods or whether it is pleasing to the gods because it is pious (1981).

That concepts have referential properties is a truism, but an important truism. A clear desideratum on a theory of concepts is that it should account for, or at least be

13. If (1) is considered to be a logical truth, then much the same point can be put by saying that the Classical Theory explains the other inferences by reducing informal validity to logical necessity.

compatible with, the referential properties of concepts.[14] According to the Classical Theory, a concept refers to those things that satisfy its definition. That is, a concept represents just those things that satisfy the conditions that its structure encodes. The appeal of this account is how nicely it meshes with the Classical Theory's other motivations. Concept acquisition, categorization, and so on are all explained in terms of the definitional structure that determines the reference of a concept. Its account of reference determination is what unifies the Classical Theory's explanatory power.

2.2. *The Retreat from Definitions*

Any theory that can do as much as the Classical Theory promises to do deserves serious consideration. In recent years, however, the theory has been subjected to intense criticism, and many feel that in spite of its obvious attractions the Classical Theory can't be made to work. We'll look at six of the main criticisms that have been raised against the Classical Theory.

Plato's Problem[15] Perhaps the most basic problem that has been leveled against the Classical Theory is that, for most concepts, there simply aren't any definitions. Definitions have proven exceptionally difficult to come by, especially if they have to be couched in perceptual or sensory terms in accordance with empiricist strictures. Locke, in discussing the concept LIE, gives a sketch of what its components should look like (1690/1975, p. 166):

> 1. Articulate Sounds. 2. Certain *Ideas* in the Mind of the Speaker. 3. Those words the signs of those *Ideas*. 4. Those signs put together by affirmation or negation, otherwise that the *Ideas* they stand for, are in the mind of the Speaker.

He adds (p. 166),

> I think I need not go any farther in the Analysis of that complex *Idea*, we call a *Lye*: What I have said is enough to shew, that it is made up of simple *Ideas*: And it could not but be an offensive tediousness to my Reader, to trouble him with a more minute enumeration of every particular simple *Idea*, that goes into this complex one; which, from what has been said, he cannot but be able to make out to himself.

Unfortunately, it is *all but obvious* how to complete the analysis, breaking the concept down into simple, sensory components. As several authors have observed (Armstrong et al. 1983 [chapter 10 in this volume]; Fodor, J. A. 1981), it isn't even clear that definitions such as the one suggested by Locke bring us any *closer* to the level of sensory

14. We say that this is a clear desideratum, but others disagree. See, e.g., Ray Jackendoff (1991) and (1989 [chapter 13 in this volume]). Jackendoff's main objection is that he thinks that reference and truth and other related notions are tied to an incorrect metaphysics, one according to which the world exists entirely independently of our ways of conceptualizing it. Jackendoff's concerns tap into deep and controversial issues in philosophy, but they are misplaced in the present context. The main distinction that we want to insist on is the difference between true and false judgments. Sometimes you are right when you think that something is a bird, sometimes you are wrong. This distinction holds whether or not *bird* is a mind-independent kind or not. To put much the same point in Kantian terms, even if we only have epistemic access to the phenomenal world, we can still make incorrect judgments about what goes on there.

15. What we call Plato's Problem shouldn't be confused with an issue which is given the same name by Noam Chomsky (1986). Chomsky's concern is with how we can know as much as we do, given our limited experience. The concern of the present section, however, is that concepts are extremely hard to define.

concepts than the concept under analysis. Are the concepts SPEAKER, AFFIRMATION, NEGATION, or STANDING FOR really any closer to the sensory level than the concept LIE.[16]

Even putting aside the empiricist strictures, however, there are few, if any, examples of definitions that are uncontroversial. Some of the most intensively studied concepts are those connected to the central topics of philosophy. Following Plato, many philosophers have tried to provide definitions for concepts like KNOWLEDGE, JUSTICE, GOODNESS, TRUTH, and BEAUTY. Though much of interest has come from these attempts, no convincing definitions have resulted.

One of the more promising candidates has been the traditional account of KNOWLEDGE as JUSTIFIED TRUE BELIEF. But even this account is now widely thought to be inadequate, in particular, because of Gettier examples (named after Edmund Gettier who first put forward an example of this kind in his 1963 paper "Is Justified True Belief Knowledge?"). Here is a sample Gettier case (Dancy 1985, p. 25):

> Henry is watching the television on a June afternoon. It is Wimbledon men's finals day, and the television shows McEnroe beating Connors; the score is two sets to none and match point to McEnroe in the third. McEnroe wins the point. Henry believes justifiably that
>
> 1 I have just seen McEnroe win this year's Wimbledon final.
>
> and reasonably infers that
>
> 2 McEnroe is this year's Wimbledon champion.
>
> Actually, however, the cameras at Wimbledon have ceased to function, and the television is showing a recording of last year's match. But while it does so McEnroe is in the process of repeating last year's slaughter. So Henry's belief 2 is true, and surely he is justified in believing 2. But we would hardly allow that Henry knows 2.

Notice that the significance of the example is that each condition in the proposed analysis of KNOWLEDGE is satisfied yet, intuitively, we all know that this isn't a case of knowledge. Philosophers concerned with the nature of KNOWLEDGE have responded in a variety of ways, usually by supplementing the analysis with further conditions (see Dancy 1985 for discussion). One thing is clear, though: Despite a tremendous amount of activity over a long period of time, no uncontroversial definition of KNOWLEDGE has emerged.

Nor is the situation confined to concepts of independent philosophical interest. Ordinary concepts have resisted attempts at definition as well. Wittgenstein (1953/1958) famously argues that the concept GAME cannot be defined. His argument consists of a series of plausible stabs at definition, followed by clear counterexamples (see the excerpt reprinted as chapter 6 in this volume). For instance, he considers and rejects the proposal that a game must be an activity that involves competition (counterexample: a card game such as patience or solitaire), or that a game must involve winning or losing (counterexample: throwing a ball against a wall and catching it).

16. A related point is that many concepts seem to involve functional elements that can't be eliminated (e.g., it may be essential to chairs that they are designed or used to be sat upon). These preclude a definition in purely sensory terms. Cf. Clark (1973), quoted above, and Miller and Johnson-Laird (1976).

In much the same spirit, Jerry Fodor (1981) considers several proposals for the concept PAINT$_{tr}$, corresponding to the transitive verb "paint." Fodor's example is quite dramatic, as he tries to show that PAINT$_{tr}$ cannot be defined even using, among other things, the concept PAINT, corresponding to the noun "paint." The first definition he considers is: X COVERS Y WITH PAINT (based on Miller 1978). Fodor argues that one reason this definition doesn't work is that it fails to provide a sufficient condition for something falling under the concept PAINT$_{tr}$. If a paint factory explodes and covers some spectators in paint, this doesn't count as an instance of PAINTING—the factory or the explosion doesn't paint the spectators—yet the case is an instance of the original proposal. What seems to be missing is that an agent needs to be involved, and the surface that gets covered in paint does so as a result of the actions of the agent. In other words: X PAINT$_{tr}$ Y if and only if X IS AN AGENT AND X COVERS THE SURFACE OF Y WITH PAINT. But this definition doesn't work either. If you, an agent, kick over a bucket of paint, and thereby cover your new shoes with paint, you haven't painted them. We seem to need that the agent intentionally covers the surface with paint. Yet even this isn't enough. As Fodor says, Michelangelo wasn't painting the ceiling of the Sistine Chapel; he was painting a picture on the ceiling. This is true, even though he was intentionally covering the ceiling with paint. The problem seems to be with Michelangelo's intention. What he primarily intended to do was paint the picture on the ceiling, not paint the ceiling. Taking this distinction into account we arrive at something like the following definition: X PAINT$_{tr}$ Y if and only if X IS AN AGENT AND X INTENTIONALLY COVERS THE SURFACE OF Y WITH PAINT AND X'S PRIMARY INTENTION IN THIS INSTANCE IS TO COVER Y WITH PAINT. Yet even this definition isn't without its problems. As Fodor notes, when Michelangelo dips his paintbrush in the paint, his primary intention is to cover the tip of his paintbrush with paint, but for all that, he isn't painting the tip of his paintbrush. At this point, Fodor has had enough, and one may have the feeling that there is no end in sight—just a boundless procession of proposed definitions and counterexamples.[17]

Of course, there could be any number of reasons for the lack of plausible definitions. One is that the project of specifying a definition is much harder than anyone has supposed. But the situation is much the same as it may have appeared to Socrates' interlocutors, as portrayed in Plato's dialogues: Proposed definitions never seem immune to counterexamples. Even the paradigmatic example of a concept with a definition (BACHELOR = UNMARRIED MAN) has been contested. Is the Pope a bachelor? Is Robinson Crusoe? Is an unmarried man with a long-term partner whom he has lived with for years?[18] As a result of such difficulties, the suspicion in much of cognitive science has come to be that definitions are hard to formulate because our concepts lack definitional structure.

17. To be fair, Fodor's discussion may not do justice to the Classical Theory. In particular, it's not clear that the force of his counterexamples stems from the meaning of PAINT$_{tr}$, rather than pragmatic factors. Certainly there is something odd about saying that Michelangelo paints his paintbrush, but the oddness may not be owing to a semantic anomaly.

18. See Fillmore (1982) and Lakoff (1987 [chapter 18 in this volume]). We should add that Lakoff's position is more complicated than just insisting that BACHELOR and the like constitute counterexamples to the Classical Theory, though others may read these cases that way. Rather, he maintains that BACHELOR has a definition but that the definition is relativized to an "idealized cognitive model" that doesn't perfectly match what we know about the world. To the extent that such mismatches occur, problematic cases arise.

The Problem of Psychological Reality A related difficulty for the Classical Theory is that, even in cases where sample definitions of concepts are granted for the purpose of argument, definitional structure seems psychologically irrelevant. The problem is that definitional structure fails to turn up in a variety of experimental contexts where one would expect it to. In particular, the relative psychological complexity of lexical concepts doesn't seem to depend on their relative definitional complexity.[19]

Consider the following example of an experiment by Walter Kintsch, which has been used to try to locate the effects of conceptual complexity in lexical concepts (reported in Kintsch 1974, pp. 230–233).[20] It is based on a phoneme-monitoring task, originally developed by D. J. Foss, where subjects are given two concurrent tasks. They are asked to listen to a sentence for comprehension and, at the same time, for the occurrence of a given phoneme. When they hear the phoneme, they are to indicate its occurrence as quickly as they can, perhaps by pressing a button. To ensure that they continue to perform both tasks and that they don't just listen for the phoneme, subjects are asked to repeat the sentence or to produce a new sentence that is related to the given sentence in some sensible way.

In Foss's original study, the critical phoneme occurred either directly after a high-frequency word or directly after a low-frequency word. He found that reaction time for identifying the phoneme correlated with the frequency of the preceding word. Phoneme detection was quicker after high-frequency words, slower after low-frequency words (Foss 1969). The natural and by now standard explanation is that a greater processing load is introduced by low-frequency words, slowing subjects' response to the critical phoneme.

Kintsch adopted this method but changed the manipulated variable from word frequency to definitional complexity. He compared subjects' reaction times for identifying the same phoneme in the same position in pairs of sentences that were alike apart from this difference: In one sentence the phoneme occurred after a word that, under typical definitional accounts, is more complex than the corresponding word in the other sentence. The stimuli were controlled for frequency, and Kintsch used a variety of nouns and verbs, including the mainstay of definitional accounts, the causatives. For example, consider the following pair of sentences:

(1) The doctor was *convinced* only by his visitor's pallor.

(2) The story was *believed* only by the most gullible listeners.[21]

This first test word ("convince") is, by hypothesis, more complex than the second ("believe"), since on most accounts the first is analyzed in terms of the second. That is, "convince" is thought to mean *cause to believe*, so that CONVINCE would have BELIEVE as a constituent.

Kintsch found that in pairs of sentences like these, the speed at which the critical phoneme is recognized is unaffected by which of the two test words precedes it. So

19. The reason the focus has been on lexical concepts is that there is little doubt that the psychological complexity associated with a phrase exceeds the psychological complexity associated with one of its constituents. In other words, the psychological reality of definitions at the level of phrases isn't in dispute.

20. For related experiments and discussion, see J. A. Fodor et al. (1980 [chapter 21 in this volume]), and J. D. Fodor et al. (1975).

21. Italics indicate the words whose relative complexity is to be tested; underlines indicate the phoneme to be detected.

the words (and corresponding concepts) that definitional accounts predict are more complex don't introduce a relatively greater processing load. The natural explanation for this fact is that definitions aren't psychologically real: The reason definitions don't affect processing is that they're not there to have any effect.

It's not obvious, however, how worried defenders of the Classical Theory ought to be. In particular, it's possible that other explanations could be offered for the failure of definitions to affect processing; definitions might be "chunked," for instance, so that they function as a processing unit. Interestingly, a rather different kind of response is available as well. Classical theorists could abandon the model of conceptual structure that these experimental investigations presuppose (viz., the Containment Model). If, instead, conceptual structure were understood along the lines of the Inferential Model, then definitional complexity wouldn't be expected to manifest itself in processing studies. The availability of an alternative model of conceptual structure shows that the experimental investigation of conceptual structure has to be more subtle. Still, Kintsch's study and others like it do underscore the lack of evidence in support of the Classical Theory. While this is by no means a decisive point against the Classical Theory, it adds to the doubts that arise from other quarters.

The Problem of Analyticity With few examples on offer and no psychological evidence for definitional structure, the burden for the Classical Theory rests firmly on its explanatory merits. We've seen that the Classical Theory is motivated partly by its ability to explain various semantic phenomena, especially analytic inferences. The present criticism aims to undercut this motivation by arguing that analyticities don't require explaining because, in fact, there aren't any. Of course, if this criticism is right, it doesn't merely challenge an isolated motivation for the Classical Theory. Rather, it calls into question the theory as a whole, since every analysis of a concept is inextricably bound to a collection of purported analyticities. Without analyticity, there is no Classical Theory.

Skepticism about analyticity is owing largely to W. V. O. Quine's famous critique of the notion in "Two Dogmas of Empiricism" [chapter 5 in this volume] and related work (see esp. Quine 1935/1976, 1954/1976). Quine's critique involves several lines of argument and constitutes a rich and detailed assessment of logical positivism, which had put analyticity at the very center of its philosophy in its distinction between meaningless pseudo-propositions and genuine (or meaningful) ones. Roughly, meaningful propositions were supposed to be the ones that were verifiable, where the meaning of a statement was to be identified with its conditions of verification. Verification, in turn, was supposed to depend upon analyticity, in that analyticities were to act as a bridge between those expressions or phrases that are removed from experience and those that directly report observable conditions. Since facts about analyticities are not themselves verifiable through observation, they needed a special epistemic status in order to be meaningful and in order for the whole program to get off the ground. The positivists' solution was to claim that analyticities are tautologies that are fixed by the conventions of a language and therefore known a priori. On this view, then, a priori linguistic analysis should be able to secure the conditions under which a statement would be verified and hence provide its meaning. This program is behind Carnap's idea that the definition or analysis of a concept provides a condition of justification for thoughts involving that concept. To be justified in thinking that

spiders are arthropods one need only verify that spiders are animals, have jointed legs, segmented bodies, and so on.

The theory that analytic statements are tautologies also helped the positivists in addressing a long-standing difficulty for empiricism, namely, how to account for the fact that people are capable of a priori knowledge of factual matters even though, according to empiricism, all knowledge is rooted in experience. Mathematics and logic, in particular, have always been stumbling blocks for empiricism. The positivists' solution was to claim that logical and mathematical statements are analytic. Since they also held that analyticities are tautologies, they were able to claim that we can know a priori the truths of logic and mathematics because, in doing so, we don't really obtain knowledge of the world (see, e.g., Ayer 1946/1952; Hahn 1933/1959).

As is clear from this brief account of the role of analyticity in logical positivism, the positivists' program was driven by epistemological considerations. The problem was, assuming broadly empiricist principles, how to explain our a priori knowledge and how to account for our ability to know and speak of scientific truths that aren't directly observable. Considering the vast range of scientific claims—that atoms are composed of protons, neutrons, and electrons, that the universe originated from a cosmic explosion 10 to 20 billion years ago, that all animals on Earth descended from a common ancestor, etc.—it is clear that the positivists' program had truly enormous scope and ambition.

Quine's attack on the notion of analyticity has several components. Perhaps the most influential strand in Quine's critique is his observation, following Pierre Duhem, that confirmation is inherently holistic, that, as he puts it, individual statements are never confirmed in isolation. As a consequence, one can't say in advance of empirical inquiry what would confirm a particular statement. This is partly because confirmation involves global properties, such as considerations of simplicity, conservatism, overall coherence, and so on. But it's also because confirmation takes place against the background of auxiliary hypotheses, and that, given the available evidence, one isn't forced to accept, or reject, a particular statement or theory so long as one is willing to make appropriate adjustments to the auxiliaries. On Quine's reading of science, no statement has an isolatable set of confirmation conditions that can be established a priori, and, in principle, there is no guarantee that any statement is immune to revision.

Some examples may help to clarify these points and ground the discussion. Consider the case of Newton's theory of gravitation, which was confirmed by a variety of disparate and (on a priori grounds) unexpected sources of evidence, such as observations of the moons of Jupiter, the phases of Venus, and the ocean tides. Similarly, part of the confirmation of Darwin's theory of evolution is owing to the development of plate tectonics, which allows for past geographical continuities between regions which today are separated by oceans. This same case illustrates the dependency of confirmation on auxiliary hypotheses. Without plate tectonics, Darwin's theory would face inexplicable data. A more striking case of dependency on auxiliary hypotheses comes from an early argument against the Copernican system that cited the absence of annual parallax of the fixed stars. Notice that for the argument to work, one has to assume that the stars are relatively close to the Earth. Change the assumption and there is no incompatibility between the Earth's movement and the failure to observe parallax. There are also more mundane cases where auxiliary hypotheses account for recalcitrant data, for instance, when college students attempt

to replicate a physical experiment only to arrive at the wrong result because of any number of interference effects. Finally, as Hilary Putnam has emphasized, a principle that appears to be immune from rejection may turn out to be one that it's rational to abandon in the context of unexpected theoretical developments. A classic example that draws from the history of science is the definition of a straight line as the shortest distance between two points—a definition that isn't correct, given that our universe isn't Euclidean. The connection between STRAIGHT LINE and THE SHORTEST DISTANCE BETWEEN TWO POINTS may have seemed as secure as any could be. Yet in the context of alternative geometries and contemporary cosmological theory, it not only turns out to be something that can be doubted, but we can now see that it is false (see Putnam 1962). What's more, Putnam and others have extended these considerations by imagining examples that illustrate the breadth of possible scientific discoveries. They've argued that we could discover, for instance, that gold or lemons aren't yellow or that cats aren't animals, thereby breaking what otherwise might have looked like the best cases of analyticities among familiar concepts.[22]

How does all this bear on the Classical Theory of concepts? Some philosophers hold that Quine has succeeded in showing that there is no tenable analytic-synthetic distinction and that this mean that concepts couldn't be definable in the way that the Classical Theory requires. However, the issue isn't so simple. Quine's critique is largely directed at the role that analyticity plays in the positivists' epistemological program, in particular, against the idea that there are statements that can be known a priori that are insulated from empirical test and that can establish specific, isolatable conditions of verification for the statements of scientific theories. If Quine is right that confirmation is holistic, then one can't establish these specific, isolatable conditions of verification. And if he is right that no statement is immune to revision, then there can't be statements that are known to be true a priori and therefore protected from future theoretical developments. So the positivist program falls flat. But the notion of analyticity needn't be tied to this explanatory burden. Analyticity simply understood as *true in virtue of meaning alone* might continue to be a viable and useful notion in describing the way that natural language and the human conceptual system works (Antony 1987; Horwich 1992). That is, for all that Quine says, there may still be a perfectly tenable analytic-synthetic distinction; it's just one that has none of the epistemological significance that the positivists took it to have. Purported analyticities are to be established on a posteriori grounds and are open to the same possibilities of disconfirmation as claims in any other part of science.

Still, Putnam's extension of Quine's considerations to examples like STRAIGHT LINE ($\neq$ SHORTEST DISTANCE ...) or GOLD ($\neq$ YELLOW METAL ...) may be disturbing to those who would like to defend the notion of analyticity. If theoretical developments allow for the rejection of these conceptual connections, then perhaps no purported analyticity will hold up to scrutiny. More or less, this direction of thought has led many philosophers to be skeptical of definitional analyses in any form, regardless of their epistemic status. The thought is that the potential revisability of nearly every statement—if only under conditions of a fantastical thought experiment—shows that the aim for definitions is futile. Yet it's hardly clear that this attitude is war-

22. For arguments that these considerations are, in fact, quite far-reaching, see Burge (1979). For arguments that we might turn out to be mistaken about the defining properties of even the paradigmatic classical concept, BACHELOR, see Lormand (1996) and Giaquinto (1996).

ranted. Its appeal may stem from paying too much attention to a limited range of examples. It may be that the cases Putnam and others have discussed are simply misleading; perhaps the concepts for the kinds in science are special. This would still leave us with thousands of other concepts. Consider, for example, the concept KILL. What surrounding facts could force one to revise the belief that killings result in death? Take someone who is honest and sincerely claims that although he killed his father, his father isn't dead or dying. No matter what the surrounding facts, isn't the plausible thing to say that the person is using the words "kill" and "dead" with anomalous meanings? At any rate, one doesn't want to prejudge cases like this on the grounds that other cases allow for revisions without changes in meaning.

In the first instance, Quine's critique of analyticity turns out to be a critique of the role of the Classical Theory in theories of justification, at least of the sort that the positivists imagined. To the extent that his arguments are relevant to the more general issue of analyticity, that's because the potential revisability of a statement shows that it isn't analytic; and many philosophers hold that this potential spans the entire language. Whether they are right, however, is an empirical question. So the issue of what analyticities there are turns on a variety of unresolved empirical matters.

The Problem of Ignorance and Error In the 1970s Saul Kripke and Hilary Putnam both advanced important arguments against *descriptivist* views of the meaning of proper names and natural kind terms (Kripke 1972/1980; Putnam 1970 [chapter 7 in this volume], 1975).[23] (Roughly, a descriptivist view is one according to which, in order to be linguistically competent with a term, one must know a description that counts as the meaning of the term and picks out its referent.) If correct, these arguments would apparently undermine the Classical Theory, which is, in effect, descriptivism applied to concepts.[24] Kripke and Putnam also sketched the outlines of an alternative positive account of the meaning of such terms, which, like their critical discussions, has been extremely influential in philosophy.

Kripke and Putnam offer at least three different types of arguments that are relevant to the evaluation of the Classical Theory. The first is an argument from error. It seems that we can possess a concept in spite of being mistaken about the properties that we take its instances to have. Consider, for example, the concept of a disease, like SMALLPOX. People used to believe that diseases like smallpox were the effects of evil spirits or divine retribution. If any physical account was offered, it was that these diseases were the result of "bad blood." Today, however, we believe that such people were totally mistaken about the nature of smallpox and other diseases. Saying this, however, presupposes that their concept, SMALLPOX, was *about* the same disease that our concept is about. They were mistaken because the disease that their concept referred to—smallpox—is very different in nature than they had supposed. Presumably, then, their most fundamental beliefs about smallpox couldn't have been part of a definition of the concept. For if they had been, then these people wouldn't have been wrong about smallpox; rather they would have been thinking and speaking

23. For arguments that similar considerations apply to an even wider range of terms, again, see Burge (1979).

24. Again, we will move freely from claims about language to claims about thought, in this case adapting Kripke's and Putnam's discussions of natural kind terms to the corresponding concepts. For an interesting discussion of how these arguments relate to the psychology of concepts, see Rey (1983 [chapter 12 in this volume]).

about some other possible ailment. Closely related to this type of argument is another, namely, an argument from ignorance. Continuing with the same example, we might add that people in the past were ignorant about a number of crucial properties of smallpox—for example, that smallpox is caused by the transmission of small organisms that multiply in great numbers inside the body of a host, and that the symptoms of the disease are the result of the causal effect of these organisms on the host's body.

Arguments from ignorance and error present compelling reasons to suppose that it's possible to possess a concept without representing necessary or sufficient conditions for its application. The conditions that a person actually associates with the concept are likely to determine the wrong extension for the concept, both by including things that do not belong in the extension, and by excluding things that do belong. By failing to represent such crucial properties of smallpox as its real nature and cause, we are likely to be left with merely symptomatic properties—properties that real cases might lack, and noncases might have.

The third type of argument is a modal argument. If an internally represented definition provides necessary and sufficient conditions for the application of a concept, it determines not just what the concept applies to as things actually stand but also what it would apply to in various possible, nonactual circumstances. The problem, however, is that the best candidates for the conditions that people ordinarily associate with a concept are ones which, by their own lights, fail to do justice to the modal facts. Thus, to change the example, we can perfectly well imagine circumstances under which gold would not have its characteristic color or other properties that we usually associate with gold. Perhaps if some new gas were to diffuse through the atmosphere, it would alter the color—and maybe various other properties—of gold. The stuff would still be gold, of course; it would simply lack its previous color. Indeed, we don't even need to imagine a hypothetical circumstance with gold, as it does lose its color and other characteristic perceptual properties in a gaseous state, yet gold-as-a-gas is still gold for all that.

One of the driving motivations behind Kripke's and Putnam's work is the intuition that we can learn important new facts about the things we think about. We can discover that gold, under other circumstances, might appear quite different to us, or that our understanding of the nature of a kind, like smallpox, was seriously in error. Discussions of these ideas are often accompanied by stories of how we might be wrong about even the most unassailable properties that are associated with ordinary concepts like GOLD, CAT, or LEMON. These stories sometimes require quite a stretch of imagination (precisely because they attempt to question properties that we would otherwise never imagine that instances of the concept could lack). The general point, however, is that we don't know which concepts we might be wrong about, or how wrong we might be. Even if some of our concepts for natural kinds have internally represented definitions which happen to determine a correct extension, it seems likely that many others do not. And if the reference of these other concepts is not mediated by definitions, we need some other account of how it is determined. This suggests that, for natural kind concepts in general, classical definitions do not mediate reference determination.

Another example might be helpful. Consider the concept HUMAN BEING. As it happens, people's views on the nature and origin of humans vary immensely. Some people believe that human beings have an immaterial soul which constitutes their

true essence. They believe that humans were created by a deity, and that they have an eternal life. Others believe that human beings are nothing but complex collections of physical particles, that they are the result of wholly physical processes, and that they have short, finite lives. And of course there are other views of humans as well.[25] Such beliefs about humans are held with deep conviction and are just the sort that one would expect to form part of a classical definition of HUMAN BEING. But presumably, at least one of these groups of people is gravely mistaken; notice that people from these different groups could—and do—argue about who is right.

How, then, is the reference of a concept to be fixed if not by an internalized definition? The Kripke/Putnam alternative was originally put forward in the context of a theory of natural language, but the picture can be extended to internal representations, with some adjustments. Their model is that a natural kind term exhibits a causal-historical relation to a kind and that the term refers to all and only members of the kind. In the present case, the assumption is that *human being* constitutes a kind and that, having introduced the term and having used it in (causal-historical) connection with humans, the term refers to all and only humans, regardless of what the people using it believe.[26]

This theory isn't without its problems, but for present purposes it pays to see how it contrasts with the Classical Theory.[27] One way to put the difference between the Kripke/Putnam account and the Classical Theory is that the Classical Theory looks to internal, psychological facts to account for reference, whereas the Kripke/Putnam account looks to external facts, especially facts about the nature of the paradigmatic examples to which a term has been historically applied. Thus much of the interest in Kripke's and Putnam's work is that it calls into question the idea that we have internally represented necessary and sufficient conditions that determine the extension of a concept.

Their arguments are similar in spirit to ones that came up in the discussion of analyticity. Here, too, classical theorists might question the scope of the objection. And, in fact, it does remain to be seen how far the Kripke/Putnam arguments for an externalist semantics can be extended. Even among the most ardent supporters of externalism, there is tremendous controversy whether the same treatment can extend beyond names and natural kind terms.

The Problem of Conceptual Fuzziness Another difficulty often raised against the Classical Theory is that many concepts appear to be "fuzzy" or inexact. For instance, Douglas Medin remarks that "the classical view implies a procedure for unambiguously determining category membership; that is, check for defining features." Yet, he adds, "there are numerous cases in which it is not clear whether an example belongs to a category" (Medin 1989, p. 1470). Are carpets furniture? One often buys carpet-

25. To mention just one, many people believe in reincarnation. Presumably, they take human beings to be something like transient stages of a life that includes stages in other organisms. It's also worth noting that past theoretical accounts of the nature of humans have been flawed. For example, neither "featherless biped" nor "rational animal" is sufficiently restrictive.

26. Michael Devitt and Kim Sterelny have done the most to develop the theory. See esp. Devitt (1981) and Devitt and Sterelny (1987).

27. The most serious of these problems has come to be known as the *Qua Problem*, that is, how to account for the fact that a word or concept has a determinate reference, despite being causally related to multiple kinds. For example, what accounts for the fact that CAT refers to cats and not to mammals, living things, or material objects? If the concept is causally related to cats, then it is automatically causally related to these other kinds too. For discussion, see Devitt and Sterelny (1987).

ing in a furniture store and installs it along with couches and chairs in the course of furnishing a home; so it may seem uncomfortable to say that carpets aren't furniture. At the same time, it may seem uncomfortable to say that they are. The problem for the Classical Theory is that it doesn't appear to allow for either indeterminacy in category membership or in our epistemic access to category membership. How can a Classical Theory account of FURNITURE allow it to be indeterminate whether carpets fall under FURNITURE, or explain how we are unable to decide whether carpets fall under FURNITURE?

Though this difficulty is sometimes thought to be nearly decisive against the Classical Theory, there are responses that a classical theorist could make. One resource is to appeal to a corresponding conceptual fuzziness in the defining concepts. Since the Classical Theory claims that concepts have definitional structure, it is part of the Classical Theory that a concept applies to all and only those things to which its definition applies. But definitions needn't themselves be perfectly sharp. They just have to specify necessary and sufficient conditions. In other words, fuzziness or vagueness needn't prohibit a definitional analysis of a concept, so long as the analysis is fuzzy or vague to exactly the same extent that the concept is (Fodor, J. A. 1975; Grandy 1990a; Margolis 1994). For instance, it is more or less uncontroversial that BLACK CAT can be defined in terms of BLACK and CAT: It is necessary and sufficient for something to fall under BLACK CAT that it fall under BLACK and CAT. All the same, we can imagine borderline cases where we aren't perfectly comfortable saying that something is or isn't a black cat (perhaps it's somewhere between determinately gray and determinately black). Admittedly, it's not perfectly clear how such a response would translate to the FURNITURE/CARPET example, but that seems more because we don't have a workable definition of either FURNITURE or CARPET than anything else. That is, the Problem of Fuzziness for these concepts may reduce to the first problem we mentioned for the Classical Theory—the lack of definitions.

The Problem of Typicality Effects The most influential argument against the Classical Theory in psychology stems from a collection of data often called *typicality effects*. In the early 1970s, a number of psychologists began studying the question of whether all instances of a given concept are on equal footing, as the Classical Theory implies. At the heart of these investigations was the finding that subjects have little difficulty ranking items with respect to how "good they are" or how "typical they are" as members of a category (Rosch 1973). So, for example, when asked to rank various fruits on a scale of 1 to 7, subjects will, without any difficulty, produce a ranking that is fairly robust. Table 1.1[28] reproduces the results of one such ranking.

What's more, rankings like these are generally thought to be reliable and aren't, for the most part, correlated with the frequency or familiarity of the test items (Rosch and Mervis 1975; Mervis, Catlin, and Rosch 1976).[29]

Typicality measures of this sort have been found to correlate with a wide variety of other psychological variables. In an influential study, Eleanor Rosch and Carolyn Mervis (1975) had subjects list properties of members of various categories. Some

28. Based on Rosch (1973), table 3. For comparison, Malt and Smith (1984) obtained the following values: Apple (6.25), Strawberry (5.0), Fig (3.38), Olive (2.25), where on their scale, 7 indicates the highest typicality ranking.

29. However, see Barsalou (1987) for a useful critical discussion of the reliability of these results.

Table 1.1

Fruit	Typicality rating on a scale of 1–7 (with 1 being highest)
Apple	1.3
Plum	2.3
Pineapple	2.3
Strawberry	2.3
Fig	4.7
Olive	6.2

Table 1.2

Feature	Bird	Robin	Chicken	Vulture
Flies	yes	yes	no	yes
Sings	yes	yes	no	no
Lays eggs	yes	yes	yes	no
Is small	yes	yes	no	no
Nests in trees	yes	yes	no	yes
Eats insects	yes	yes	no	no

properties occurred in many of the lists that went with a category, others occurred less frequently. What Rosch and Mervis found was that independent measures of typicality predict the distribution of properties that occur in such lists. An exemplar is judged to be typical to the extent that its properties are held to be common among other exemplars of the same superordinate category.[30] For instance, robins are taken to have many of the properties that other birds are taken to have, and correspondingly, robins are judged to be highly typical birds, whereas chickens or vultures, which are judged to be significantly less typical birds, are taken to have fewer properties in common with other birds (see table 1.2).[31]

Importantly, typicality has a direct effect on categorization when speed is an issue. The finding has been, if subjects are asked to judge whether an X is a Y, that independent measures of typicality predict the speed of correct affirmatives. So subjects are quicker in their correct responses to "Is an apple a fruit?" than to "Is a pomegranate a fruit?" (Rosch 1973; Smith, Shoben, and Rips 1974). What's more, error rates correlate with typicality. The more typical the probe relative to the target category, the fewer errors.[32]

The problem these results pose for the Classical Theory is that it has no natural model for why they should occur. Rather, the Classical Theory seems to predict that

30. In the literature, *exemplar* is used to denote subordinate concepts or categories, whereas *instance* is used to denote individual members of a given category.

31. Based on Smith (1995), table 1.3.

32. Typicality measures correlate with a variety of other phenomena as well. See Rosch (1978 [chapter 8 in this volume]).

all exemplars should be on a par. If falling under BIRD is a matter of satisfying some set of necessary and sufficient conditions, then all (and only) birds should do this equally. And if categorizing something as a bird is a matter of determining that it satisfies each of the required features for being a bird, there is no reason to think that "more typical" exemplars should be categorized more efficiently. It's not even clear how to make sense of the initial task of rating exemplars in terms of "how good an example" they are. After all, shouldn't all exemplars be equally good examples, given the Classical Theory's commitment that they all satisfy the same necessary and sufficient conditions for category membership?

In an important and influential overview of the intellectual shift away from the Classical Theory, Edward Smith and Douglas Medin note that there are, in fact, classical models that are compatible with various typicality results (Smith and Medin 1981). As an example, they suggest that if we assume that less typical members have more features than typical ones, and we also assume that categorization involves an exhaustive, serial, feature-matching process, then less typical members should take longer to categorize and cause more processing errors. After all, with more features to check, there will be more stages of processing. But the trouble with this and related models is that they involve ad hoc assumptions and conflict with other data. For instance, there is no reason to suppose that atypical exemplars have more features than typical ones.[33] Also, the model incorrectly predicts that atypical exemplars should take longer to process in cases where the categorization involves a negated target (an *X* is not a *Y*). It should take longer, that is, to judge that a chicken is not a fish than to judge that a robin is not a fish, but this just isn't so. Finally, the account has no explanation of why typicality correlates with the distribution of features among exemplars of a superordinate category.

Also, it's worth noting that the features that are involved in the typicality data are not legitimate classical features since most are not necessary. A quick look at table 1.2 makes this clear: *none* of the features listed there is necessary for being a bird; none is shared by all three exemplars. So an explanation in terms of the number of features can't really get off the ground in the first place, since the features at stake aren't classical.

In sum, then, typicality effects raise serious explanatory problems for the Classical Theory. At the very least, they undermine the role of the Classical Theory in categorization processes. But, more generally, they suggest that the Classical Theory has little role to play in explaining a wide range of important psychological data.

The Classical Theory has dominated theorizing about concepts from ancient times until only quite recently. As we have just seen, though, the theory is not without serious problems. The threats posed by these objections are not all of the same strength, and, as we've tried to emphasize, the Classical Theory has some potential responses to mitigate the damage. But the cumulative weight against the theory is substantial and has been enough to make most theorists think that, in spite of its impressive motivations, the Classical Theory simply can't be made to work.

33. If anything, it would be the opposite, since subjects usually list more features for typical exemplars than for atypical ones. But one has to be careful about taking "feature lists" at face value, as the features that subjects list are likely to be governed by pragmatic factors. For instance, no one lists for BIRD that birds are objects. Most likely this is because it's so obvious that it doesn't seem relevant.

Box 2

Summary of Criticisms of the Classical Theory

1. **Plato's Problem**
   There are few, if any, examples of defined concepts.
2. **The Problem of Psychological Reality**
   Lexical concepts show no effects of definitional structure in psychological experiments.
3. **The Problem of Analyticity**
   Philosophical arguments against analyticity also work against the claim that concepts have definitions.
4. **The Problem of Ignorance and Error**
   It is possible to have a concept in spite of massive ignorance and/or error, so concept possession can't be a matter of knowing a definition.
5. **The Problem of Conceptual Fuzziness**
   The Classical Theory implies that concepts have determinate extensions and that categorization judgments should also yield determinate answers, yet concepts and categorization both admit of a certain amount of indeterminacy.
6. **The Problem of Typicality Effects**
   Typicality effects can't be accommodated by classical models.

3. *The Prototype Theory of Concepts*

3.1. *The Emergence of Prototype Theory*

During the 1970s, a new view of concepts emerged, providing the first serious alternative to the Classical Theory. This new view—which we will call the *Prototype Theory*—was developed, to a large extent, to accommodate the psychological data that had proved to be so damaging to the Classical Theory. It was the attractiveness of this new view, as much as anything else, that brought about the downfall of the Classical Theory.

There is, of course, no single account to which all prototype theorists subscribe. What we are calling the Prototype Theory is an idealized version of a broad class of theories, which abstracts from many differences of detail. But once again putting qualifications to the side, the core idea can be stated plainly. According to the Prototype Theory, most concepts—including most lexical concepts—are complex representations whose structure encodes a statistical analysis of the properties their members tend to have.[34] Although the items in the extension of a concept *tend* to have these properties, for any given feature and the property it expresses, there may be items in the extension of a concept that fail to instantiate the property. Thus the features of a concept aren't taken to be necessary as they were on the Classical Theory. In addition, where the Classical Theory characterized sufficient conditions for concept application in terms of the satisfaction of all of a concept's features, on the Prototype Theory application is a matter of satisfying a sufficient number of features, where some may be weighted more significantly than others. For instance, if BIRD is composed of such features as FLIES, SINGS, NESTS IN TREES, LAYS EGGS, and so on, then on the

34. More likely they are structured and interconnected sets of features (Malt and Smith 1984). For example, with the concept BIRD, features for size and communication might be linked by the information that small birds sing and large birds don't.

Prototype Theory, robins are in the extension of BIRD because they tend to have all of the corresponding properties: robins fly, they lay eggs, etc. However, BIRD also applies to ostriches because even though ostriches don't have all of these properties, they have enough of them.[35]

This rejection of the Classical Theory's proposed necessary and sufficient conditions bears an affinity to Wittgenstein's suggestion that the things that fall under a concept often exhibit a family resemblance. They form "a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail" (Wittgenstein 1953/1968 [chapter 6 in this volume], p. 32). In fact, Eleanor Rosch and Carolyn Mervis, two important and influential figures in the development of the Prototype Theory, explicitly draw the parallel to Wittgenstein's work (1975, p. 603):

> The present study is an empirical confirmation of Wittgenstein's (1953) argument that formal criteria are neither a logical nor psychological necessity; the categorical relationship in categories which do not appear to possess criterial attributes, such as those used in the present study, can be understood in terms of the principle of family resemblance.

For Wittgenstein, as for Rosch and Mervis, a word or concept like GAME isn't governed by a definition but rather by a possibly open-ended set of properties which may occur in different arrangements. Some games have these properties, some have those, but despite this variation, the properties of games overlap in a way that establishes a similarity space. What makes something a game is that it falls within the boundaries of this space.

Because the Prototype Theory relaxes the constraints that the Classical Theory imposes on a concept's features, it is immune to some of the difficulties that are especially challenging for the Classical Theory. First among these is the lack of definitions. Since the Prototype Theory claims that concepts don't have definitional structure, it not only avoids but actually predicts the difficulty that classical theorists have had in trying to specify definitions. Similarly, the Prototype Theory is immune to the problems that the Classical Theory has with analyticity. Given its rejection of the classical idea that concepts encode necessary conditions for their application, the Prototype Theory can wholeheartedly embrace the Quinean critique of analyticity. Additionally, the theory makes sense of the fact that subjects generally list non-necessary properties in the generation of feature lists.

The rejection of necessary conditions also highlights the Prototype Theory's emphasis on nondemonstrative inference. This is, in fact, another advantage of the theory, since one function of concepts is to allow people to bring to bear relevant information upon categorizing an instance or exemplar. Yet encoding information isn't without its tradeoffs. As Rosch puts it, "[T]he task of category systems is to provide maximum information with the least cognitive effort...." (1978 [chapter 8 in this volume], p. 28). What this means is that representational systems have to strike

35. For convenience, it will be useful to refer to a such structure as a concept's "prototype." We should point out, however, that the term "prototype" doesn't have a fixed meaning in the present literature and that it's often used to refer to the exemplar that has the highest typicality ratings for a superordinate concept (as, e.g., when someone says that ROBIN is the prototype for BIRD).

a balance.[36] On the one hand, a concept should encode a considerable amount of information about its instances and exemplars, but on the other, it shouldn't include so much that the concept becomes unwieldy. The solution offered by the Prototype Theory is that a concept should encode the distribution of statistically prominent properties in a category. By representing statistically prominent properties, concepts with prototype structure generate many more inferences than do classical representations; they trade a few maximally reliable inferences for many highly reliable though fallible ones.[37]

The Prototype Theory also has an attractive model of concept acquisition—in fact, much the same model as the Classical Theory. In both cases, one acquires a concept by assembling its features. And, in both cases, it's often assumed that the features correspond to sensory properties. The main difference is that on the Prototype Theory, the features of a concept express statistically prominent properties. So on the Prototype Theory the mechanism of acquisition embodies a statistical procedure. It doesn't aim to monitor whether various properties always co-occur, but only whether they tend to. Of course, to the extent that the Prototype Theory inherits the empiricist program associated with the Classical Theory, it too faces the problem that most concepts resist analysis in sensory terms. The trouble with empiricism, remember, isn't a commitment to definitions but a commitment to analyzing concepts in purely sensory terms. If LIE was a problem for Locke, it's just as much a problem for prototype theorists. Assuming they can articulate some plausible candidate features, there is still no reason to think that all of these can be reduced to a sensory level. This is true even for their stock examples of concepts for concrete kinds, concepts like BIRD or FRUIT.[38] But, like the Classical Theory, the Prototype Theory can be relieved of its empiricist roots. When it is, its model of concept acquisition is at least as compelling as the Classical Theory's.

Probably the most attractive aspect of the Prototype Theory is its treatment of categorization. Generally speaking, prototype theorists model categorization as a similarity comparison process that involves operations on two representations—one for the target category and one for an instance or an exemplar. (For ease of expression, we'll frame the discussion in terms of instances only, but the same points go for exemplars as well.) On these models, an instance is taken to be a member of a category just in case the representation of the instance and the representation of the category are judged to be sufficiently similar. The advantage of this approach is that similarity-based categorization processes lay the groundwork for a natural explana-

36. Rosch, however, sharply distances herself from any psychological interpretation of this work (see Rosch 1978). But as we are interested in the bearing of research in this tradition on theories of concepts construed as mental particulars, we will not discuss nonpsychological interpretations.

37. For Rosch, much of the interest in the efficiency of a conceptual system concerns its hierarchical structure. "[N]ot all possible levels of categorization are equally good or useful; rather, the most basic level of categorization will be the most inclusive (abstract) level at which the categories can mirror the structure of attributes perceived in the world" (1978, p. 30). According to Rosch and her colleagues the basic level in a conceptual system is defined in terms of its informational potential relative to other levels in the hierarchy, and its effects are widespread and can be independently measured. For instance, basic level concepts appear early in cognitive and linguistic development, they have priority in perceptual categorization, and, in a hierarchy, they pick out the most abstract categories whose members are similar in shape. For discussion, see Rosch (1978) and Rosch et al. (1976).

38. Look at most discussions and you'll find that the sample features for BIRD are things like WINGS, FLIES, EATS WORMS, SINGS, and so on. Notice, though, that none of these is more "sensory" than BIRD itself.

tion of typicality effects. To see how this works, we need to take a closer look at the notion of similarity.

Prototype theorists have developed a number of different psychological measures for similarity. Perhaps the most commonly used is Amos Tversky's (1977) "Contrast Principle" (see, e.g., Smith et al. 1988 [chapter 17 in this volume]).[39] The idea behind this principle is that the judged similarity of any two items, *i* and *j*, is measured by comparing the sets of shared and distinctive features that are associated with them. Where I and J are the feature sets, the function can be defined as follows:

$$\text{Sim}\ (I, J) = af(I \cap J) - bf(I - J) - cf(J - I)$$

The constants a, b, and c allow for different weights to be assigned to the set of common features ($I \cap J$) and to each set of distinctive features ($I - J$ and $J - I$), and the function *f* allows for weights to be assigned to individual features. To illustrate how the principle works, consider the measure of similarity between BIRD and TWEETIE, where the latter is a representation that, for simplicity, incorporates just four features: FLIES, SINGS, IS SMALL, and LAYS EGGS. Also assume that the sets of common and distinctive features are each given an equal weight of 1 (i.e., a, b, and c are all 1) and that the function *f* assigns each of the individual features equal weight. Then, using the six features in table 1.2, the similarity of TWEETIE to BIRD is $4 - 2 - 0 = 2$. Presumably, this is sufficiently high to count Tweetie as a bird.[40]

Now the Contrast Principle measures the psychological similarity of two categories, but it doesn't specify the computational procedure that actually generates the judgment. For a sample processing model, consider this simple schematic account (see Smith and Medin 1981; Smith 1995): To compute the similarity of a given object to a target category, one compares the feature sets associated with the object and the category, possibly checking all the features in parallel. As each feature is checked, one adds a positive or negative value to an accumulator, depending on whether it is a common feature or not. When the accumulator reaches a certain value, the judgment is made that the item is sufficiently similar to the target category to count as a member; items that are computed to have a lower value are judged insufficiently similar—they are taken to be nonmembers.

This isn't the only model of categorization that is open to prototype theorists. Yet even one as straightforward as this generates much of the typicality data:

> *Graded Judgments of Exemplariness* Recall the datum that subjects find it a natural task to rank exemplars for how typical they are for a given category. Apples are judged to be more typical of fruit than olives are. The accumulator model explains this phenomena under the assumption that the very same mechanism that is responsible for categorization is also responsible for typicality judgments. Since the mechanism results in a similarity judgment, and since similarity is itself a graded notion, it's no surprise that some exemplars are considered to be more typical than others. The ones that are more similar to the

39. For other measures of similarity, see Shepard (1974) and Estes (1994). For further discussion, see Medin et al. (1993), Gleitman et al. (1996), and *Cognition* 65, nos. 2–3—a special issue devoted to the topic of similarity.

40. The same measure also works in the comparison of a representation of an exemplar and a superordinate concept. For instance, using table 1.2 again, the similarity of ROBIN to BIRD is $6 - 0 - 0 = 6$, and the similarity of CHICKEN to BIRD is $1 - 5 - 0 = -4$.

target are the ones that are judged to be more typical; the ones that are less similar to the target are the ones that are judged to be less typical.

Typicality Correlates with Property Lists The reason the distribution of features in subjects' property lists predicts the typicality of an exemplar is that the properties that are the most common on such lists characterize the structure of the concept that is the target of the similarity-comparison process. Taking the example of BIRD and its exemplars, the idea is that the properties that are commonly cited across categories such as *robin, sparrow, hawk, ostrich,* and so on, are the very properties that correspond to the features of BIRD. Since ROBIN has many of the same features, robins are judged to be highly typical birds. OSTRICH, on the other hand, has few of these features, so ostriches are judged to be less typical birds.

Graded Speed of Quick Categorization Judgments Assuming that the individual feature comparisons in the similarity-comparison process take varying amounts of time, the outcome of each comparison will affect the accumulator at different times. As a result, items that are represented to have more features in common with a target will be judged more quickly to be members. A less thorough comparison is required before a sufficient number of shared features is registered.

Categorization Errors Are Inversely Correlated with Typicality For less typical exemplars, more feature comparisons will be needed before a sufficient number of shared features is reached, so there are more chances for error.

The accumulator model also explains certain aspects of conceptual fuzziness. Prototype theorists often cite fuzziness as a point in favor of their theory, while not saying much about what the fuzziness of concepts consists in. One way of unpacking the notion, however, is that judgments about whether something falls under a concept are indeterminate, that is, the psychological mechanisms of categorization do not yield a judgment one way or the other.

Fuzziness To predict fuzziness in this sense, the model need only be supplemented with the following qualification: Where an exemplar isn't clearly similar enough to a target by a prespecified margin the result is neither the judgment that it falls under the target concept nor the judgment that it doesn't.

From this brief survey of the data, one can see why the Prototype Theory has been held in such high regard. Not only does it seem to be immune to some of the difficulties surrounding the Classical Theory, but it addresses a wide variety of empirical data as well. While there is virtually no doubt about the importance of these data, a number of problems have been raised for the theory, problems that are largely directed at its scope and interpretation. Some of these problems have been thought to be serious enough to warrant a radical reworking of the theory, or even its abandonment. We'll discuss four.

Box 3

The Prototype Theory

Most concepts (esp. lexical concepts) are structured mental representations that encode the properties that objects in their extension tend to possess.

3.2. Problems for the Prototype Theory

The Problem of Prototypical Primes In an important early critical discussion of the Prototype Theory, Sharon Armstrong, Lila Gleitman, and Henry Gleitman investigated the question of whether well-defined concepts, such as EVEN NUMBER or GRANDMOTHER, exhibit typicality effects (Armstrong et al. 1983). ("Well-defined" here means that people know and can readily produce the concepts' definitions.) Armstrong et al. argued that if typicality effects reveal that a concept has statistical structure, then well-defined concepts shouldn't exhibit typicality effects. Using four well-defined concepts, they showed that people nonetheless find it natural to rank exemplars according to how good they are as members of such concepts.[41] Just as apples are ranked as better examples of fruit than figs are, the number 8 is ranked as a better example of an even number than the number 34 is. What's more, Armstrong et al. found that typicality rankings for well-defined concepts correlate with other data in accordance with some of the standard typicality effects. In particular, typicality correlates with speed and accuracy of categorization. Just as subjects produce correct answers for "Is an apple a fruit?" faster than for "Is a fig a fruit?" they produce correct answers for "Is 8 an even number?" faster than for "Is 34 an even number?" The conclusion that Armstrong et al. reached was that the considerations that are standardly thought to favor the Prototype Theory are flawed. "[T]o the extent that it is secure beyond doubt that, e.g., FRUIT and PLANE GEOMETRY FIGURE have different structures, a paradigm that cannot distinguish between responses to them is not revealing about the structure of concepts" (p. 280). In other words, Armstrong et al. took their findings to be evidence that typicality effects don't argue for prototype structure.

A common way of thinking about prototypes—and the one that Armstrong et al. assume—is to interpret a concept with prototype structure as implying that subjects represent its extension as being graded. On this view of prototypes, subjects think that robins are literally "birdier" than ostriches, just as Michael Jordan is literally taller than Woody Allen. The reason prototypes are read this way is because of the focus on typicality judgments. Typicality judgments are then explained as reflecting people's views about the degree to which the instances of an exemplar instantiate a category. Unsatisfied with the argument that moves from typicality judgments to prototype structure, Armstrong et al. asked subjects outright whether various categories are graded, including their four well-defined categories. What they found was that, when asked directly, people actually claim that well-defined concepts aren't graded—and many hold that other categories, such as *fruit*, aren't graded either—but even so they remain willing to rank exemplars for how good they are as members. Although Armstrong et al.'s subjects unanimously said that *even number* is an all-or-none category, the tendency was still to say 8 is a better example of an even number than 34 is.

Armstrong et al. took this to be further evidence that the arguments for prototype structure involve deep methodological problems. Yet this may be too strong of a conclusion. One could hold instead that typicality effects do argue for prototype

41. The four concepts Armstrong et al. investigated were EVEN NUMBER, ODD NUMBER, FEMALE, and PLANE GEOMETRY FIGURE. Though they didn't test the concept PRIME NUMBER, we feel it's safe to say that this concept would exhibit the same effects. For example, we bet that subjects would say that 7 is a better example of a prime number than 113 is.

structure but that prototype structure has no implications for whether subjects represent a category as being graded. In other words, the proposal is that typicality judgments reflect an underlying prototype; it's just that prototypes needn't involve a commitment to graded membership.

If typicality judgments aren't about degrees of membership, what are they about? We are not sure that there is a simple answer. Yet it's not unreasonable to think much of what's going on here relates back to properties that are represented as being highly indicative of a category. The difference between ROBIN and OSTRICH, on this view, is that robins are represented as possessing more of the properties that, for one reason or another, are taken to be the usual signs that something is a bird. But the usual signs needn't themselves be taken to be constitutive of the category. So long as one believes that they aren't, and that they merely provide evidence for whether something is a member of the category, the number of signs an item exhibits needn't determine a degree to which it instantiates the category.

The distinction between properties that are represented as being evidential and those that are represented as being constitutive is especially pertinent when categorization takes place under pressures of time and limited resources. In a pinch, it makes sense to base a categorization judgment on the most salient and accessible properties—the very ones that are most likely to be merely evidential. The conclusion that many psychologists have drawn from this observation is that categorization can't be expected to be a univocal affair. Given the correlations between judged typicality and quick category judgments for both accuracy and speed, the Prototype Theory provides a compelling account of at least part of what goes on in categorization. But considered judgments of category membership seem to tell a different story. This has prompted a variety of theorists to put forward *Dual Theories* of concepts, where one component (the "identification procedure") is responsible for quick categorization judgments and the other component (the "core") is called upon when cognitive resources aren't limited (Osherson and Smith 1981 [chapter 11 in this volume]; Smith et al. 1984; Landau 1982).[42] Such Dual Theories have often been thought to give the best of both worlds—the Prototype Theory's account of fast categorization and the Classical Theory's account of more thoughtful categorization, especially where the relevant properties are hidden or in some way less accessible. For instance, in discussing the merits of Dual Theories, Smith et al. (1984) are careful to insist that both the core and the identification procedure are accessed in categorization processes. The difference between them, they claim, can be illustrated with the concept GENDER. "Identification properties might include style of clothing, hair, voice, etc., while core properties might involve having a particular kind of sexual organs. As this example suggests, our distinction centers on notions like salience, computability, and diagnosticity..." (p. 267).

42. The division of labor between the core and the identification procedure hasn't been fully worked out in the literature. For instance, in the text we adopt the interpretation according to which the difference between cores and identification procedures is just a matter of how they enter into categorization processes. Another difference that's often cited is that cores are the primary, or perhaps the only, component that enters into the compositional principles that determine the semantics of complex concepts on the basis of their constituents. But it is at least open to question whether the components responsible for making considered judgments of category membership are also the ones that compositionally generate the semantics of complex concepts. We discuss this issue further below.

Unfortunately, such a view ignores the difficulties that are associated with the theories it tries to combine. For instance, if there was a problem before about specifying definitions, adding a prototype component to a classical component doesn't eliminate the problem. Nor does it help with the Problem of Ignorance and Error, which, as it turns out, arises for both theories in isolation and so can't help but arise for a Dual Theory.

The Problem of Ignorance and Error Since the Prototype Theory requires a way of fixing the extensions of concepts, ignorance and error are still as much a problem as they were for the Classical Theory. Indeed, in some ways they are actually more of a problem for the Prototype Theory. Take, for example, the concept GRANDMOTHER. Prototypical grandmothers are old, they have gray hair and glasses, they are kind to children, and, let's suppose, they like to bake cookies. The problem is that someone can satisfy these properties without being a grandmother, and someone can be a grandmother without satisfying these properties. Tina Turner is a grandmother. So is Whoopi Goldberg.

Much the same point applies to concepts that lack definitions or whose definitions aren't generally known. Consider, once again, the concept SMALLPOX. The properties that most people associate with this disease, if any, are its symptoms—high fever, skin eruptions, and so on. And since symptoms are, in general, reliable effects of a disease, they are good candidates for being encoded in prototype representations. At the same time, the Prototype Theory faces a serious difficulty: Because symptoms aren't constitutive of a disease but are instead the effects of a variety of causal interactions, they aren't completely reliable guides to the presence of the disease. Someone could have the symptoms without having the disease, and someone could have the disease without the symptoms. As Armstrong et al. note, birds with all their feathers plucked are still birds, and "3-legged, tame, toothless, albino tigers" are still tigers (1983, p. 296). Nor is a convincing toy tiger a tiger. The point is that everyone knows this and is prepared to acknowledge it, so, by their own lights, prototype representations don't determine the correct extension for a concept like BIRD or TIGER. Prototype representations lack sufficient richness to include all birds or all tigers, and at the same time they are, in a sense, too rich in that they embody information that includes things that aren't birds or tigers.

One way to avoid these conclusions that some might find tempting is to claim that if something doesn't fit a concept's prototype, then it doesn't really fall under the concept. That is, one might make the radical move of denying that TIGER applies to our toothless, 3-legged creature. The idea behind this suggestion is that how a concept is deployed determines what items fall under it. Yet while this view may have some initial appeal, it can't be made to work—it's really far too crude. Not only would it imply that 3-legged albino tigers aren't tigers and that convincing tiger toys are, but in general, it would rule out the possibility of *any* misrepresentation. When Jane is nervously trekking through the Amazon jungle, fearful of snakes, and she is startled by what she takes to be a snake lying across her path just ahead, we want it to be possible that she could actually be mistaken, that it could turn out that she was startled by a snake-shaped vine, and not a snake at all. But if categorization processes determine the extension of the concept, then this item has to be a snake: Since it was categorized as falling under SNAKE, it *is* a snake. In short, on this suggestion there is no

room for the possibility of a concept being misapplied, and this is just too high a price to pay.[43]

Notice that Dual Theories might help somewhat, if it's assumed that conceptual cores are involved in categorization. The core would provide Jane with a definition of SNAKE that would have the final word on whether something falls under the concept by providing a more substantial procedure for deciding whether something is a snake. Then her mistake could be credited to the deployment of an identification procedure; what would make it a mistake is that the outcome of the identification procedure fails to match the outcome of the core. Presumably, were Jane to deploy the core, she'd be in a position to recognize her own error. But as we've already noted, Dual Theories aren't much of an advance, since they reintroduce the difficulties that face the Classical Theory.

Another mark against the present form of a Dual Theory is that it inherits the difficulties associated with a verificationist semantics. For instance, people's procedures for deciding whether something falls under a concept are subject to change as they acquire new information, new theories, and (sometimes) new technologies. Yet this doesn't mean that the concept's identity automatically changes. To return to the example of a disease, when two people differ on the symptoms they associate with measles, they would appear to be in disagreement; that is, they appear to be arguing about the best evidence for deciding whether measles is present. But if the identity of MEASLES is given by the procedures under which one decides whether it is instantiated, then we'd have to say that the two couldn't genuinely disagree about the symptoms associated with measles. At best, they would be talking at cross purposes, one about one ailment, the other about another. The same goes for a single person over time. She couldn't come to change her mind about the best indications of measles, since in adopting a new procedure of verification she'd thereby come to deploy a new concept. We take it that these difficulties offer good prima facie grounds for shying away from a verificationist version of the Dual Theory.

The Missing Prototypes Problem The strongest evidence in favor of the Prototype Theory is that subjects find it natural to rate exemplars and instances in terms of how representative they are of a given category and the fact that these ratings correlate with a range of psychological phenomena. But although this is true of many concepts, it is by no means true of all concepts. Many concepts aren't associated with typicality judgments, and for many concepts, people fail to represent any central tendencies at all. As Jerry Fodor has put it (1981, pp. 296–297):

> There may be prototypical *cities* (London, Athens, Rome, New York); there may even be prototypical *American cities* (New York, Chicago, Los Angeles), but there are surely no prototypical *American cities situated on the East Coast just a little south of Tennessee*. Similarly, there may be prototypical *grandmothers* (Mary Worth) and there may be prototypical *properties of grandmothers* (*good, old* Mary Worth). But there are surely no prototypical properties of, say, *Chaucer's grand-*

43. Note that nothing turns on the example being a natural kind (where it's plausible that science is the best arbiter of category membership). The point is just that, wherever there is representation, there is the potential for misrepresentation. An account that doesn't permit misrepresentation simply isn't an adequate theory of concepts.

mothers, and there are no prototypical properties of *grandmothers most of whose grandchildren are married to dentists.*

It's important to see that this is not at all an isolated problem, or an artifact of a few exotic examples. Indefinitely many complex concepts lack prototype structure. Some fail to have prototype structure because people simply don't have views about the central tendencies of the corresponding categories. This seems to be the case with many uninstantiated concepts:

- U.S. MONARCH
- 4TH CENTURY SAXOPHONE QUARTET
- 31ST CENTURY INVENTION
- GREAT-GREAT-GREAT GRANDCHILD OF CINDY CRAWFORD

Others lack prototype structure because their extensions are too heterogeneous:

- A CONSEQUENCE OF PHYSICAL PROCESSES STILL GOING ON IN THE UNIVERSE
- OBJECTS THAT WEIGH MORE THAN A GRAM
- NEW SPECIES
- NOT A WOLF
- FROG OR LAMP

Still others lack prototype structure for other reasons:

- BELIEF[44]
- THE RADIATION BEING THE SAME IN EVERY DIRECTION TO A PRECISION OF ONE PART IN ONE HUNDRED THOUSAND
- PIECE OF PAPER I LEFT ON MY DESK LAST NIGHT
- IF X IS A CHAIR, X IS A WINDSOR[45]

A related problem is that it's perfectly possible to have a concept without knowing a prototype, even if others who possess the concept do. Thus, for example, you could have the concept of a DON DELILLO BOOK or a FRISBEE-GOLF COACH without representing any properties as being statistically prominent in the corresponding categories, even though other people may have strong views about the matter. Delillo fans know that his books are usually funny, they have slim plots, and are laced with poignant observations of American popular culture. But if you haven't read a Delillo book, you may not know any of this. Still, what's to stop you from possessing the concept, using it to support inductive inferences, organize memory, or engage in categorization? If you know that Don Delillo's books are usually well stocked at Barnes and Noble, then you may infer that Barnes and Noble is likely to have Delillo's latest book. If you are told that his latest is *Underworld*, then you will remember it as a Delillo book. And later, when you go to Barnes and Noble and you see a copy of *Underworld*, you will categorize it as a Delillo book. It would seem, then, that concept possession doesn't require a representation with prototype structure.

44. Osherson and Smith (1981) suggest that concepts like BELIEF, DESIRE, and JUSTICE may lack prototype structure because they are too "intricate"—a somewhat vague yet intriguing idea.

45. For some discussion of concepts that involve Boolean constructions, see Fodor (1998). Fodor points out that these concepts are generally subject to what he calls the *Uncat Problem*, namely, they lack prototypes.

The objection that many concepts lack prototype structure is standardly presented as an issue about compositionality, since most of the concepts that lack prototypes are patently complex. Compositionality is certainly an important feature of the conceptual system, as it provides the best explanation for one of the most important and striking features of human thought—its productivity. Important as compositionality is, however, it's not really needed for the present objection. The force of the Missing Prototypes Problem is simply that many concepts lack prototype structure and that it's often possible to possess a concept without thereby knowing a prototype.

The implications of this objection aren't always given their full due. Edward Smith, for example, suggests that the Prototype Theory isn't intended to be a general theory of concepts. He says that some classes, such as *objects that weigh forty pounds*, are arbitrary and that "the inductive potential of a class may determine whether it is treated as a category" (1995, p. 7). The representation OBJECTS THAT WEIGH FORTY POUNDS, however, is a perfectly fine concept, which one can readily use to pick out a property. For any of a variety of purposes, one might seek to find an object that weighs forty pounds, categorize it as such, and reason in accordance with the corresponding concept. In any event, though there is nothing wrong with the idea that concepts divide into groups requiring different theoretical treatments, we still require an account of the concepts that aren't covered by the Prototype Theory. Given that there seem to be indefinitely many such concepts, the question arises whether prototypes are central and important enough to concepts generally to be considered part of their nature. Perhaps it is more appropriate to say that many lexical concepts have prototypes associated with them but that these prototypes aren't in any way constitutive of the concepts.

Another option—one that aims to mitigate the damage caused by the Missing Prototypes Problem—is (once again) to appeal to a Dual Theory. The idea might be that for some concepts it is possible to have the concept without having both components. So for these concepts, not knowing a prototype is fine. The advantage of this sort of Dual Theory would appear to be that it allows for a univocal treatment of all concepts; one needn't appeal to a completely distinct theory for those concepts that lack prototypes. Yet it's hardly clear that this is much of a gain, since the resulting Dual Theory fails to preserve the spirit of the Prototype Theory. It looks like what's essential to a concept, on this view, is the classical core, with the prototype being (in many cases) merely an added option. In short, the Dual Theory is beginning to sound more and more like a supplemented version of the Classical Theory.

The Problem of Compositionality One of the most serious and widely discussed objections to the Prototype Theory is the charge that it's unable to account for the phenomenon of compositionality. This difficulty seems especially pressing in light of the importance of compositionality in accounting for our ability to entertain an unbounded number of concepts. To the extent that anyone can foresee an explanation of this ability, it's that the conceptual system is compositional.[46]

Early discussions of compositionality in the literature on Prototype Theory were concerned with explaining how graded extensions could be combined. Thus these discussions were based on the assumption that most categories are graded in the

46. Which isn't to say that the details have been completely worked out or that there is no controversy about the content of the principle of compositionality. For discussion, see Grandy (1990b).

sense that items are members of a category to varying degrees (i.e., membership isn't an all-or-none matter).[47] The standard model for composing graded categories was a version of fuzzy set theory—a modification of standard set theory that builds on the notion of graded membership (see esp. Zadeh 1965). A fuzzy set can be understood in terms of a function that assigns to each item in the domain of discourse a number between 0 and 1, measuring the degree to which the item is in the set. If an item is assigned the value 1, it is wholly and completely inside the set. If it is assigned the value 0, it is wholly and completely outside the set. All values between 0 and 1 indicate intermediate degrees of membership, with higher values indicating higher degrees. Under these assumptions, fuzzy set theory characterizes a variety of operations that are analogues of the standard set-theoretic operations of intersection, union, and so on. Fuzzy set intersection, for example, is given in terms of the *Min Rule*: An item is a member of the fuzzy intersection of two sets to the minimum of the degrees to which it is an element of the two sets. If Felix is a cat to degree 0.9 and is ferocious to degree 0.8, then Felix is a ferocious cat to degree 0.8.[48]

In a seminal discussion of the Prototype Theory's reliance on fuzzy sets, Daniel Osherson and Edward Smith presented a number of forceful objections to this treatment of compositionality (Osherson and Smith 1981). One is a straightforward counterexample to the Min Rule. Consider the intersective concept STRIPED APPLE (intersective in that intuitively its extension is determined by the intersection of the corresponding categories—something is a striped apple just in case it's striped and an apple). Fuzzy set theory reconstructs this intuition by saying that the concept's extension is determined by fuzzy set intersection. That is, something is a striped apple to the minimum of the degrees that it is striped and that it is an apple. A consequence of this view is that nothing should be counted as a striped apple to a higher degree than it is counted as an apple. But, as Osherson and Smith point out, a very good instance of a striped apple will inevitably be a poor instance of an apple. The Min Rule simply makes the wrong prediction.[49] Perhaps more worrying still, consider the concept APPLE THAT IS NOT AN APPLE. Clearly, the extension of this concept is empty; it's logically impossible for something that is not an apple to be an apple. Yet fuzzy set theory's account of compositionality doesn't deliver this result. APPLE THAT IS NOT AN APPLE is just another intersective concept, combining APPLE and NOT AN APPLE. According to the Min Rule, something falls under it to the minimum of the degrees to which it is an apple and to which it is not an apple. Taking again a highly representative striped apple, we may suppose that such an item is taken to be an apple to a fairly low degree (perhaps 0.3) and striped to some higher degree (perhaps 0.8). Taking the complement of the fuzzy set of apples, our item is not-an-apple to the degree $1 - 0.3 = 0.7$. Since it will be an instance of APPLE THAT IS NOT AN APPLE to the minimum of the degrees to which it is an instance of APPLE (0.3) and to which it is an instance of NOT AN APPLE (0.7), it will be an instance of APPLE THAT IS NOT AN APPLE to degree 0.3.

47. This assumption seemed plausible to many in light of the fact that subjects were so willing to rate instances or exemplars of a concept in terms of how representative they were of the concept. But again, the results of Armstrong et al. (1983) show that the inference from such ratings to graded membership is mistaken.

48. In like fashion, the complement of a fuzzy set may be defined by taking the value of $1 - x$ for each element of the set. E.g., if Felix is in the set of cats to degree 0.9, then Felix is in the set of non-cats to degree $1 - 0.9 = 0.1$.

49. For an argument against a broader class of proposals (of which the Min Rule is a special case), see Osherson and Smith (1982).

Though difficulties like these may seem to be decisive against fuzzy set theory's model of compositionality, we should note that fuzzy set theory doesn't provide the only model of compositionality that is compatible with the Prototype Theory.[50] Still, compositionality has proven to be a notable stumbling block for prototypes.

The general objection that Prototype Theory cannot provide an adequate account of conceptual combination has been pushed most vigorously by Jerry Fodor. In this context, Fodor has argued both that many complex concepts simply don't have prototypes and that, when they do, their prototypes aren't always a function of the prototypes of their constituents. We've already dealt with the first sort of case, under the heading of the Problem of Missing Prototypes. To get a feel for the second, consider the concept PET FISH. The prototype for PET FISH is a set of features that picks out something like a goldfish. Prototypical pet fish are small, brightly colored, and they live in fish bowls (or small tanks). How does the prototype for PET FISH relate to the prototypes of its constituents, namely, PET and FISH?[51] Presumably, the features that constitute the prototypes for PET pick out dogs and cats as the most representative examples of pets—features such as FURRY, AFFECTIONATE, TAIL-WAGGING, and so on. The prototype for FISH, on the other hand, picks out something more like a trout or a bass—features such as GRAY, UNDOMESTICATED, MEDIUM-SIZED, and so on. Thus prototypical pet fish make rather poor examples both of pets and of fish. As a result, it's difficult to see how the prototype of the complex concept could be a function of the prototypes of its constituents.

One of the most interesting attempts to deal with the composition of complex prototypes is Smith, Osherson, Rips, and Keane's (1988 [chapter 17 in this volume]) Selective Modification Model. According to this model, conceptual combinations that consist of an adjectival concept (e.g., RED, ROUND) and a nominal concept (e.g., APPLE, FRUIT) in the form Adj + N are formed by a process where the adjectival concept modifies certain aspects of the nominal concept's structure. The nominal concept is taken to decompose into a set of features organized around a number of attributes. Each attribute is weighted for diagnosticity, and instead of having default values, each value is assigned a certain number of "votes," indicating its probability. For simplicity, Smith et al. consider only adjectival concepts assumed to have a single attribute (see figure 1.1). The way conceptual combination works is that the adjectival concept selects the corresponding attribute in the nominal concept's representation, increases its diagnosticity, and shifts all of the votes within the scope of the attribute to the value that the adjectival concept picks out. For instance, in the combination RED APPLE, the attribute COLOR is selected in the representation APPLE, its diagnosticity is increased, and the votes for all of the color features are shifted to RED (see figure 1.2).

Smith et al. subjected this model to the following sort of experimental test. By asking subjects to list properties of selected items, they obtained an independent measure of the attributes and values of a range of fruit and vegetable concepts. They took the number of listings of a given feature to be a measure of its salience (i.e., its number of votes), and they measured an attribute's diagnosticity by determining how useful it is in distinguishing fruits and vegetables. This allowed them to generate

50. Indeed, Osherson and Smith have proposed an alternative model of their own, which we will discuss shortly. See also Hampton (1991).

51. We take it that the empirical claims made here about the prototypes of various concepts are extremely plausible in light of other findings, but the claims are not based on actual experimental results. Accordingly, the arguments ultimately stand in need of empirical confirmation.

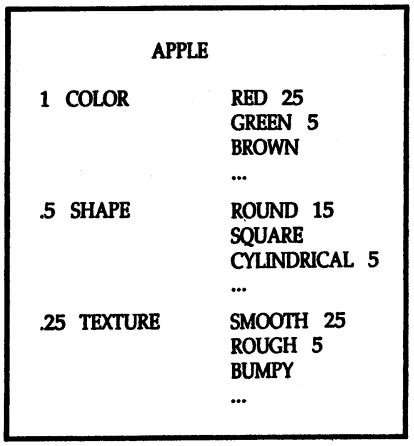

Figure 1.1
A partial representation of the structure of the concept APPLE. Each attribute (COLOR, SHAPE, TEXTURE) is weighted for diagnosticity, represented by the number to the left of the attribute. The values (RED, GREEN, ROUND, etc.) are each assigned a certain number of "votes," indicating their probability.

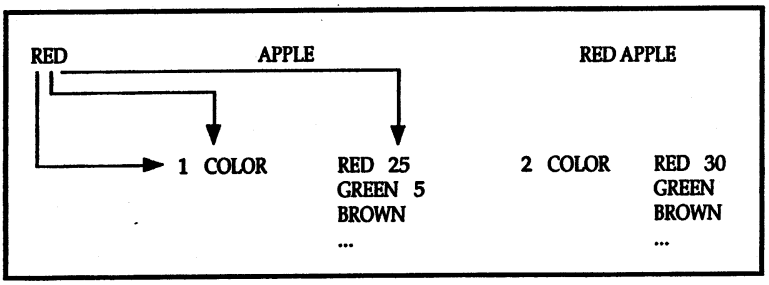

Figure 1.2
A schematic representation of the Selective Modification Model. RED combines with APPLE by selecting the attribute COLOR, increasing its diagnosticity, and shifting all of votes within its scope to RED (adapted from Smith et al. 1988, p. 493).

predictions for the typicality of exemplars for complex concepts such as RED VEGETABLE, ROUND FRUIT, and LONG VEGETABLE. Then they compared these predictions with the typicality ratings that subjects gave in an independent test. The average of the correlations between predictions and directly elicited typicality ratings was 0.70.[52]

Despite this success, the Selective Modification Model is highly limited—a point that Smith et al. themselves bring attention to. Even if we restrict the scope of a compositional theory to the simplest sorts of complex concepts, it doesn't cover nonintersective concepts (e.g., FAKE, ALLEGED, POSSIBLE) and it is especially unequipped to deal with cases where the modifier's effects transcend a single attribute, as, for example, with the concept WOODEN SPOON. (Wooden spoons are known to be larger

52. For vegetable concepts the average was 0.88. Abstracting from a few anomalous results which may have been due to a poor choice of exemplars, the average of all correlations would have been 0.87. See Smith et al. (1988) for details and further tests of the model.

than other spoons and used for cooking, not eating.)[53] It doesn't even cover the case we started with, namely, PET FISH.

Smith et al. suggest some ways in which the model might attempt to cope with these difficulties. One borrows an idea from James Hampton (1987), who notes that the prototypes for some complex concepts may be sensitive to real-world knowledge. For instance, your prototype for WOODEN SPOON may be more a result of experience with wooden spoons than your having constructed the concept from compositional principles. In Smith et al.'s hands, this suggestion emerges as a two-stage model. In the first stage, a prototype is constructed on a purely compositional basis, in accordance with the original mechanism of the Selective Modification Model; in the second, the prototype is subject to changes as world knowledge is brought into play. In principle, a more complicated model like this is capable of dealing with a fair number of the difficult examples we've mentioned. For instance, WOODEN SPOON needn't be so troublesome anymore. Perhaps people do construct a prototype in which just the MATERIAL COMPOSITION attribute for spoon is altered. Later, in the second stage, the attribute SIZE is altered as experience teaches that wooden spoons are typically larger than metal spoons. Perhaps PET FISH can be accommodated by a two-stage model as well.

The strongest objection to Hampton's suggestion is owing to Jerry Fodor and Ernest Lepore. They emphasize that one can't allow experience to fix the prototype of a complex concept without admitting that such prototypes are essentially idioms. But, they argue, if prototypes are idioms, then the Prototype Theory offers a wholly inadequate account of concepts (1996, p. 267):

> Prototypes aren't compositional; they work like idioms. Concepts, however, must be compositional; nothing else could explain why they are productive. So concepts aren't prototypes.

In addition, they argue that the two-stage model is implausible since as concepts get more complex (and we are less likely to have real-world knowledge about them), we don't default to a compositionally determined prototype. As an example, they point to the concept PET FISH WHO LIVE IN ARMENIA AND HAVE RECENTLY SWALLOWED THEIR OWNERS. Though no one has real-world knowledge for a concept like this—knowledge that might interfere with the effects of the Selective Modification Model—no one has a compositionally determined prototype either. Concepts like these simply lack prototypes.

Notice that the second of these objections is no more than a repetition of the Missing Prototypes Problem. The reply here will be much the same as it was there.[54]

53. For another example, consider MALE NURSE. Male nurses aren't taken to be just like other nurses, only male. Among other things, they wear different sorts of uniforms—slacks, not dresses. Thus the combination can't just be a matter of the modifier affecting the SEX attribute in NURSE, shifting all the votes to the value MALE. For some discussion of the significance of context effects in conceptual combination, see Medin and Shoben (1988).

54. Actually, we aren't so sure that highly modified concepts inevitably lack prototypes. For many cases it seems likely that people will have a sketchy idea of how to rank exemplars or instances for typicality. Take Fodor and Lepore's example: While we aren't prepared to say too much about these unusual fish, we do know they have to be fairly large if they are going to swallow people (who but a person owns a fish?). Among other things, this knowledge implies that goldfish are going to be extremely poor exemplars and that white sharks may be better. To the extent that one can make such judgments, this counts as evidence for a schematic prototype. If it's idiomatic, that's just to say that there are other ways to construct an idiomatic prototype than by having experience with members of the corresponding category. In this case, the idiom could derive from a reasoning process that incorporates information from the classical core and general background knowledge.

Smith et al. are free to adopt a Dual Theory.[55] Under a Dual Theory, concepts have two components—a classical core and an optional identification procedure with prototype structure. Since one can possess a concept while being in possession of just a core, the absence of a prototype is no problem at all. What's more, the absence of a prototype needn't prevent the concept from being compositional, so long as the core is compositional. And everyone agrees that if compositionality works anywhere, it works for classical conceptual components. In short, the failure of prototypes to compose doesn't argue against the Prototype Theory once it's admitted that concepts aren't *just* prototypes. Fodor and Lepore's arguments have no leverage against a Dual Theory.

On the other hand, we will need an account of how prototypes are constructed for those complex concepts that do have prototypes. Since people can generate prototypes for some novel complex concepts in the absence of any specific experience with members of the corresponding category, the implication is that at least part of the story will be compositional (cf. STRIPED APPLE, WOODEN BICYCLE, ORANGE ELEPHANT). This is the context in which the Selective Modification Model should be viewed. To the extent that compositional processes are responsible for the construction of prototypes, the model is pertinent. What the model doesn't aim to do is provide a comprehensive account of the composition of concepts. A theory of prototype composition is one thing, a theory of concept composition is another. Under a Dual Theory, concepts aren't (just) prototypes.

Fodor (1998) has another argument against Smith et al., but it too falls short once the implications of a Dual Theory are recognized. His argument is that PET FISH couldn't be an idiom, since it clearly licenses the inferences PET FISH → PET and PET FISH → FISH. In contrast, a paradigmatic idiom like KICKED THE BUCKET doesn't generate any such inferences. (If John kicked the bucket, it doesn't follow that there is something he kicked.) Fodor's gloss of this contrast is that the inferences in the case of PET FISH result from the compositional structure of the concept, in particular, its logical form. But, he claims, under the Prototype Theory, concepts don't have logical forms; they have prototype structure. By now the flaw in this reasoning should be fairly clear. A prototype theorist who opts for a Dual Theory can claim that concepts *do* have logical forms insofar as they have classical cores. PET FISH needn't be an idiom, even if its prototype is.[56]

Still, relying on a Dual Theory isn't unproblematic. Our main worry for the Prototype Theory in connection with the Smith et al. model of prototype combination is that prototypes seem more and more like cognitive structures that are merely associated with concepts rather than structures that are part of the nature of concepts. The more it's granted that prototypes are optional and that the prototypes for complex concepts act like idioms, the less essential prototypes seem to be. Once again,

55. Though they aren't perfectly explicit about the matter, it appears that they do adopt a Dual Theory when they claim that "prototypes do not exhaust the contents of a concept" (Smith et al. 1988, p. 486).

56. For this same reason, it won't do for Fodor and Lepore to argue that the weights assigned to the modified features aren't compositionally determined. "[W]hat really sets the weight of PURPLE in PURPLE APPLE isn't its prototype; it's its logical form" (1996, p. 264). Fodor and Lepore's point is that the modified feature in a simple construction is often given maximum weight, as if it didn't express a statistical property at all. True enough, but this needn't be anything more than a reflection of PURPLE APPLE's classical core. Alternatively, Smith et al. could add that their second stage of processing has access to the concept's core, letting classical modifiers adjust the corresponding features so that they receive a maximum weight.

with the core of a concept apparently doing so much work, the Dual theory beings to look more like a supplemented version of the Classical Theory. We should end this discussion, however, by emphasizing that the issues surrounding compositionality are extremely complicated and that there is much more to be said. We'll return to these issues in what we hope will be a new and illuminating context, when we examine some of the problems associated with Conceptual Atomism (sec. 6.2).

The Prototype Theory continues to be one of the dominant theories of concepts in psychology and cognitive science. This is understandable, given its ability to explain a wide range of psychological data. We've seen, however, that in the face of a number of problems related to concept possession and reference determination, prototype theorists are apt to fall back on the idea that concepts have classical cores. The result is that the Prototype Theory may inherit some of the difficulties that motivated it in the first place. This may be so, regardless of how strong the evidence is that concepts have prototype structure.

Box 4

Summary of Criticisms of the Prototype Theory

1. **The Problem of Prototypical Primes**
   Typicality effects don't argue for prototype structure, since even well-defined concepts exhibit typicality effects.
2. **The Problem of Ignorance and Error**
   Ignorance and error is as much a problem for the Prototype Theory as it is for the Classical Theory. Indeed, the problem is considerably worse for the Prototype Theory, since concepts with prototype structure fail to cover highly atypical instances and incorrectly include non-instances.
3. **The Missing Prototypes Problem**
   Many concepts lack prototypes.
4. **The Problem of Compositionality**
   The Prototype Theory does not have an adequate account of compositionality, since the prototypes of complex concepts aren't generally a function of the prototypes of their constituent concepts.

4. *The Theory-Theory of Concepts*

4.1. *Theories, Explanations, and Conceptual Structure*

In the past ten years or so, an increasing number of psychologists have gravitated to a view in which cognition generally is assimilated to scientific reasoning. The analogy to science has many strands. One is to distance the theory of categorization from early empiricist models, where categorization consisted of nothing more than a process of checking an instance against a list of sensory properties. Another is to liken concepts to theoretical terms, so that philosophical treatments of theoretical terms can be recruited in psychology. Yet another is to provide a characterization and explanation of conceptual change along the lines of theory change in science. Within the boundaries of these explanatory goals lies the *Theory-Theory of Concepts.*[57]

57. The terminology here is somewhat unfortunate, since "Theory-Theory" is also used in reference to a specific account of how people are able to attribute mental states to one another. The view is that they have an internalized theory of mind. See, e.g., Wellman (1990).

People who approach the Theory-Theory for the first time may find it somewhat confusing, because theory-theorists slip between talking about concepts being like theories and concepts being like theoretical terms—structures at entirely different levels. When theory-theorists say that concepts are mental theories, using expressions like the child's or the adult's "theory of number," the intended object of investigation is a body of propositions that articulate people's knowledge within a given domain. When theory-theorists say that concepts are like theoretical terms, they are concerned with the constituents of thoughts. The trouble, of course, is that the Theory-Theory can't at once be about concepts understood in both of these ways; that would amount to a mereological paradox.

A natural bridge between these two ways of appealing to theories is to give priority to the second notion (where concepts are likened to theoretical terms) but to explain their nature relative to the first notion. Susan Carey holds a view like this. The focus of much of Carey's research has been the characterization of how children understand things differently than adults in several important domains of cognition. In laying out the background to her investigations she is unusually explicit in isolating concepts from larger cognitive structures (Carey 1991 [chapter 20 in this volume], p. 258):

> Concepts are the constituents of beliefs; that is, propositions are represented by structures of concepts. Theories are complex mental structures consisting of a mentally represented domain of phenomena and explanatory principles that account for them.

And in her seminal book *Conceptual Change in Childhood*, she draws the connection between concepts and the mental theories in which they are embedded (1985, p. 198):

> One solution to the problem of identifying the same concepts over successive conceptual systems and of individuating concepts is to analyze them relative to the theories in which they are embedded. Concepts must be identified by the roles they play in theories.

In other words, the idea is that some bodies of knowledge have characteristics that distinguish them as analogues to scientific theories and that the concepts that occur in these bodies of knowledge are individuated by their cognitive roles in their respective "mental theories."

This view raises a number of questions, one of which is whether any cognitive structures warrant the designation of mental theory. Among theory-theorists, there is considerable disagreement about how lenient one should be in construing a body of representations as a theory. Most would agree that an important feature of theories is that they are used for explanatory purposes.[58] Yet this alone doesn't help, since it just raises the issue of how permissive one should be in treating something as an appropriate explanation. Carey, for one, is fairly restrictive, claiming that only a dozen or so cognitive structures should be counted as theories (1985, p. 201). On the other side of the spectrum, Gregory Murphy and Douglas Medin are so permissive that they count nearly any body of knowledge as a theory (1985 [chapter 19 in this volume]).[59] We

58. For this reason, the Theory-Theory is sometimes called the *Explanation-Based View* (see, e.g., Komatsu 1992).

59. "When we argue that concepts are organized by theories, we use *theory* to mean any of a host of mental 'explanations,' rather than a complete, organized, scientific account. For example, causal knowledge certainly embodies a theory of certain phenomena; scripts may contain an implicit theory of the entailment relations between mundane events; knowledge of rules embodies a theory of the relations between rule constituents; and book-learned, scientific knowledge certainly contains theories" (Murphy and Medin, 1985, 290).

don't want to have to settle this dispute here, so we'll opt for a more permissive understanding of theories. For our purposes, the point to focus on is that a concept's identity is determined by its role within a theory.

Now there would be little to argue about if the claim were merely that concepts are embedded in explanatory schemas of sorts. Few would deny this. The interesting claim is that a concept's identity is constituted by its role in an explanatory schema. To put this claim in a way that brings out its relation to other theories of concepts, we can say that according to the Theory-Theory concepts are structured mental representations and that their structure consists in their relations to other concepts specified by their embedding theories. Notice that put this way the Theory-Theory can't appeal to the Containment Model of conceptual structure. For any two concepts that participate in the same mental theory, the structure of each will include the other; but if the first contains the second, the second can't contain the first. What this shows is that the Theory-Theory is partial to the Inferential Model of structure. Concepts are individuated in virtue of the inferences they license based on their role in the theories that embed them.

When it comes to concept application, the Theory-Theory appeals to the structure of a concept, just as the Classical Theory and the Prototype Theory do. Generally, psychologists haven't been explicit about how the mechanism works, but their remarks about how they view scientific terms places them squarely in a tradition that is familiar from the philosophy of science (see, e.g., Kuhn 1962; Sellars 1956; and Lewis 1970, 1972). On this account the meaning of a theoretical term is determined by its role in a scientific theory. This can be given as a definite description that characterizes the role that the term plays in the theory.[60] Then the referent of the term is whatever unique entity or kind satisfies the description.[61]

One advantage of the Theory-Theory is in the models of categorization that it encourages. Many psychologists have expressed dissatisfaction with earlier theories of concepts on the grounds that they fail to incorporate people's tendency toward essentialist thinking—a view that Douglas Medin and Andrew Ortony (1989) have dubbed *psychological essentialism*. According to psychological essentialism, people are apt to view category membership for some kinds as being less a matter of an instance's exhibiting certain observable properties than the item's having an appropriate internal structure or some other hidden property. For instance, we all recognize the humor in the Warner Brothers cartoons involving Pepe LePew. In these sketches, a delicate and innocent black female cat is subjected to the inappropriate attention of a gregarious male skunk when she accidentally finds herself covered head to toe by a stripe of white paint. The joke, of course, is that she isn't a skunk, even though to all appearances she looks like one. As most people see it, what makes something a skunk isn't the black coat and white markings, but rather having the right biological history, or the right genetic make-up.

It's not just adults who think this. Prompted by an interest in the development of essentialist thinking, a number of psychologists have investigated its emergence in

60. See Lewis's papers, in particular, for an account based on the work of Frank Ramsey (1929/1990) which shows how one can provide definite descriptions for theoretical terms when their meanings are interdefined.

61. An alternative account, which theory-theorists generally haven't explored, is to say that much of the content of a concept is given by its role in cognition but that its referent is determined independently, perhaps by a causal relation that concepts bear to items in the world. Cf. two-factor conceptual role theories in philosophy, such as Block (1986).

childhood. Susan Gelman and Henry Wellman, for instance, have found marks of psychological essentialism in children as young as four and five years old (Gelman and Wellman 1991 [chapter 26 in this volume]).[62] Young children, it turns out, are reasonably good at answering questions about whether a substantial transformation of the insides or outsides of an object affects its identity and function. When asked if an item such as a dog that has had its blood and bones removed is still a dog, Gelman and Wellman's young subjects responded 72% of the time that it no longer is. And when asked whether the same sorts of items change identity when their outsides are removed (in this case, the dog's fur), they responded 65% of the time that they do not.

The Theory-Theory connects with psychological essentialism by allowing that people access a mentally represented theory when they confront certain category decisions. Rather than passing quickly over a check-list of properties, people ask whether the item has the right hidden property.[63] This isn't to say that the Theory-Theory requires that people have a detailed understanding of genetics and chemistry. They needn't even have clearly developed views about the specific nature of the property. As Medin and Ortony put it, people may have little more than an "essence placeholder" (1989, p. 184). We gather that what this means is that people represent different sorts of information when they think of a kind as having an essence. In some cases they may have detailed views about the essence. In most, they will have a schematic view, for instance, the belief that genetic makeup is what matters, even if they don't represent particular genetic properties or have access to much in the way of genetic knowledge.

Earlier, in looking at the Prototype Theory, we saw that categorization isn't necessarily a single, unitary phenomenon. The mechanisms responsible for quick categorization judgments may be quite different from the ones responsible for more considered judgments. If anything, the Theory-Theory is responsive to people's more considered judgments. This suggests that a natural way of elaborating the Theory-Theory is as a version of the Dual Theory. As before, the identification procedure would have prototype structure, only now, instead of a classical core, concepts would have cores in line with the Theory-Theory. We suspect that a model of this sort has widespread support in psychology.

Apart from its ties to categorization, much of the attraction of the Theory-Theory has come from its bearing on issues of cognitive development. One source of interest in the Theory-Theory is that it may illuminate the cognitive differences between children and adults. In those cases where children have rather different ways of conceptualizing things than adults, such a difference may be due to children and adults' possessing qualitatively distinct theories. Cognitive development, on this view, mimics the monumental shifts in theories that are exhibited in the history of science (Carey 1985, 1991; Keil 1989; Gopnik and Meltzoff 1997). Some theorists would even go further, arguing that theory changes in development are due to the very same cognitive mechanisms that are responsible for theory change in science. On this view, the claim isn't merely that an analogy exists between scientists and children; the claim is rather that scientists and children constitute a psychological kind. As

62. See also Carey (1985), Keil (1989), and Gelman et al. (1994).

63. As a result, the Theory-Theory, like the Prototype Theory, is concerned with nondemonstrative inference. In conceptualizing an item as falling under a concept, the inferences that are licensed include all of those that go with thinking of it as having an essence. For example, in categorizing something as a bird, one is thereby licensed to infer that it has whatever essence is represented for birds and that its salient observable properties (e.g., its wings, beak, and so on) are a causal effect of its having this essence.

Alison Gopnik puts it, "Scientists and children both employ the same particularly powerful and flexible set of cognitive devices. These devices enable scientists and children to develop genuinely new knowledge about the world around them" (1996, p. 486; see also Gopnik and Meltzoff 1997). In other words, cognitive development and theory change (in science) are to be understood as two facets of the very same phenomenon.

In sum, the Theory-Theory appears to have a number of important advantages. By holding that concepts are individuated by their roles in mental theories, theory-theorists can tie their account of concepts to a realistic theory of categorization—one that respects people's tendency toward essentialist thinking. They also can address a variety of developmental concerns, characterizing cognitive development in terms of the principles relating to theory change in science. Despite these attractions, however, the Theory-Theory isn't without problems. Some shouldn't be too surprising, since they've cropped up before in other guises. Yet the Theory-Theory also raises some new and interesting challenges for theorizing about concepts.

Box 5

The Theory-Theory

Concepts are representations whose structure consists in their relations to other concepts as specified by a mental theory.

4.2. Problems for the Theory-Theory

The Problem of Ignorance and Error Let's start with the Problem of Ignorance and Error. Does it affect the Theory-Theory too? It certainly does, and in several ways. For starters, we've seen that theory-theorists typically allow that people can have rather sketchy theories, where the "essence placeholder" for a concept includes relatively little information. Notice, however, that once this is granted, most concepts are going to encode inadequate information to pick out a correct and determinate extension. If people don't represent an essence for birds, apart from some thin ideas about genetic endowment, then the same goes for dogs, and bears, and antelopes. In each case, the theory in which the concept is embedded looks about the same. People have the idea that these creatures have some property in virtue of which they fall into their respective categories, but they don't have much to say about what the property is. How, then, will these concepts come to pick out their respective extensions?

When we faced a comparable problem in the context of the Prototype Theory, the natural solution was to rely on a Dual Theory that posited classical cores. If prototypes don't determine reference (because of the Problem of Ignorance and Error), then perhaps that isn't their job; perhaps they should be relegated to identification procedures. Within the context of the Theory-Theory, however, the analogous move is something of a strain. As we've noted, the Theory-Theory is generally understood to be about considered acts of categorization and hence is itself most naturally construed as giving the structure of conceptual cores. In any event, it's not likely that appealing to the Classical Theory can help, since it too faces the Problem of Ignorance and Error.

A lack of represented information isn't the only difficulty for the Theory-Theory. In other cases, the problem is that people represent incorrect information. A simple

example is that someone might incorporate a false belief or two into their essence placeholder for a concept. To return to our example from before, someone might hold that smallpox is caused by divine retribution. But, again, this shouldn't stop him from entertaining the concept SMALLPOX, that is, the very same concept that we use to pick out a kind that has nothing in particular to do with God. To the extent that Putnam and Kripke are right that we might be incorrect in our deeply held beliefs about a kind, the same point holds for the Theory-Theory.[64]

To take another example, consider people's concept PHYSICAL OBJECT. Elizabeth Spelke, Renée Baillargeon, and others have tried to characterize this concept, while engaging in a sustained and fascinating program of research which asks whether infants have it too (see, e.g., Spelke 1990; Baillargeon 1993 [chapter 25 in this volume]; Leslie 1994; and Gopnik and Meltzoff 1997). Generally speaking, the notion of a physical object that has emerged is one of a cohesive three-dimensional entity that retains its boundaries and connectedness over time. Among the principles that are widely thought to underlie people's understanding of such things is that qua physical objects, they can't act upon one another at a distance.[65] For example, were a moving billiard ball to come close to a stationary ball yet stop just short of touching it, one wouldn't view the subsequent movement of the stationary ball as being a causal effect of the first ball's motion, even if it continued in the same direction as the first ball. This principle—sometimes called the *principle of contact*—seems to encapsulate deeply held beliefs about physical objects, beliefs that can be traced back to infancy.

Notice, however, that the principle of contact is in direct conflict with physical principles that we all learn in the classroom. The first billiard ball may not crash into the other, but it still exerts a gravitational influence on it, however small. The implication is that most people's understanding of physical objects may be in error. The very entities that people are referring to in thinking about physical objects lack a property that is about as fundamental to their understanding of physical objects as one can imagine. In other words, their theory of physical objects is incorrect, yet this doesn't stop them from thinking about physical objects. Of course, one could try to maintain the stark position that prior to being educated in the science of physics such people aren't wrong about anything. They simply have a different concept than the rest of us. This position might be explored in more detail, but we don't think it's especially attractive. The reason is, once again, that one wants to say that these people could change their minds about the nature of objects or that they could be in a position of arguing with their educated counterparts. To the extent that such disagreements are possible, the concepts that are pitted against one another have to be in some sense the same. Otherwise, there wouldn't be any disagreement—just a verbal dispute.

The Problem of Stability To be sure, whether two people are employing the same concept or not and whether the same person is employing the same concept over time are difficult questions. For purposes of setting out the Problem of Ignorance and Error, we've relied on a number of cases where intuitively the same concept is at play. We suspect, however, that many theorists would claim that it's simply inappropriate to insist that the very same concept may occur despite a difference in surrounding

64. Thus it's ironic that discussions of the Theory-Theory sometimes take it to be a development of Kripke's and Putnam's insights about natural kind terms.

65. The qualification is to preclude cases of psychological action at a distance. That is, objects understood as psychological entities may cause each other to move without being in contact with one another, but objects understood as purely physical bodies cannot.

beliefs. The alternative suggestion is that people need only have similar concepts. That is, the suggestion is to concede that differences in belief yield distinct concepts but to maintain that two concepts might be similar enough in content that they would be subsumed by the same psychological generalizations.

Suppose, for instance, that your theory of animals says that animals are entirely physical entities while your friend's theory of animals says that some animals (perhaps humans) have nonphysical souls. This might mean that you don't both possess the same concept ANIMAL. Still, by hypothesis, you both possess concepts with similar contents, and though strictly speaking they aren't the same, they are similar enough to say that they are both animal-concepts. Let's call the problem of explaining how the content of a concept can remain invariant across changes in belief, or how two people with different belief systems can have concepts with the same or similar content, the *Problem of Stability*. The suggestion that is implicit in many psychological discussions is that strict content stability is a misguided goal. Really what matters is content similarity. As Smith et al. (1984) put it, "[T]here is another sense of stability, which can be equated with similarity of mental contents (e.g., 'interpersonal stability' in this sense refers to situations where two people can be judged to have similar mental contents) ..." (p. 268).

As tempting as this strategy may be, it's not as easy to maintain as one might have thought. The difficulty is that the notion of content similarity is usually unpacked in a way that presupposes a prior notion of content identity (Fodor and Lepore 1992). Consider, for instance, Smith et al.'s explanation. They propose that two concepts are similar in content when they have a sufficient number of the same features. Moreover, they point out that subjects tend to cite the same properties in experiments where they are asked to list characteristics of a category. Following Rosch and others, they take this to be evidence that people's concepts, by and large, do incorporate the same features. The consequence is supposed to be that people's concepts are highly similar in content.

But notice the structure of the argument. Features are themselves contentful representations; they are just more concepts. Smith et al.'s reasoning, then, is that two concepts are similar in content when their structure implicates a sufficient number of concepts with the *same* content. But if these other concepts have to share the same content, then that's to say that the notion of content similarity is building upon the notion of content identity; the very notion that content similarity is supposed to replace is hidden in the explanation of how two concepts could be similar in content. What's more, Smith et al.'s proposal is hardly idiosyncratic. Content similarity is generally understood in terms of overlapping sets of features. But again, feature sets can't overlap unless they have a certain number of the same features, that is, representations with the same content. And if they have representations with the same content, then one might as well admit that concepts have to have the same content (not similar content), despite differences in belief. This brings us full circle.

The scope of this problem hasn't been absorbed in the cognitive science community, so perhaps it pays to consider another proposed solution. Here's one owing to Lance Rips (1995). He suggests that we think of concepts as being individuated along two dimensions. One is a mental theory; the other, a formally specified mental symbol. So the concept DOG is a formally individuated mental representation taken together with a collection of contentful states that incorporate salient information about dogs. Rips likens his model to a Dual Theory of concepts, but one that incorporates neither a classical core nor a prototype-based identification procedure. The advantage of the model is supposed to be that without postulating definitions for

concepts, Rips's "cores" provide sufficient resources to solve a number of problems, including the problem of stability. They are supposed to generate stability, since states can be added or removed from the theory part of a concept while the core remains invariant. In this way, changes or differences in belief can still be tracked by the same mental representation. Consequently, there is a mechanism for saying that they are changes, or differences, with respect to the same theory.

Now Rips himself admits that his account doesn't have a fully developed explanation of stability. Yet he claims to have solved the problem for cases where the belief changes are relatively small (Rips 1995, p. 84):

> To take the extreme case, if there is *no* overlap in your previous and subsequent theories of daisies then does your former belief that Daisies cause hayfever conflict with your present belief that *Daisies don't cause hayfever*? The present proposal leaves it open whether a larger divergence in representations about a category [i.e., the theory component of a concept] could force a change in the representations-of the category [i.e., the formally individuated symbol]. What's clear is that less drastic differences in a theory do allow disagreements, which is what the present suggestion seeks to explain.

In other words, changes in a small number of the beliefs that make up a given theory needn't undermine stability, so long as the subsequent theory is associated with the very same formally identified symbol.

This is a novel and interesting suggestion, but unfortunately it can't be made to work as it stands. The reason is that incidental changes to a theory can't be tracked by a representation understood as a merely formal item. That's like tracking the content of a cluster of sentences by reference to a word form that appears throughout the cluster. Notice that whether the cluster of sentences continues to mean the same thing (or much the same thing) depends upon whether the invariant word form continues to mean the same thing (or much the same thing). If for some reason the word comes to have a completely different content, then the sentences would inherit this difference. If, for example, the word form starts out by expressing the property *electron* but later comes to express the property *ice cream*, the subsequent theory wouldn't conflict with the previous theory. In short, Rips's suggestion doesn't get us very far unless his "core" part of the concept, that is, the symbol, maintains its content over time. Then one could easily refer back to the content of that symbol in order to claim that the earlier theory and the subsequent theory are both about electrons. But Rips can't accept this amendment; it assumes that a concept's content is stable across changes in belief. Rather than explain stability, it presupposes stability.[66]

This isn't the last word on conceptual stability. We expect that other suggestions will emerge once the issue is given more attention. Nonetheless, stability is one of the key problems that a worked-out version of the Theory-Theory needs to face.[67]

66. Another way to make the main point here is to ask what makes something a *small* change in a theory. Intuitively, small changes are ones that don't affect the contents of the concepts involved, and Rips seems to be saying just that. His story amounts to the claim that concepts are stable (i.e., they don't change meaning) under relatively small changes in theories (i.e., changes that don't affect meaning). Clearly, without an independent account of when a change is small, this theory is vacuous.

67. That there are few discussions of stability is, we think, a reflection of the fact that the Theory-Theory hasn't been subjected to as much critical scrutiny as previous theories. Another respect in which the Theory-Theory remains relatively undeveloped is in its treatment of compositionality. On the face of it, theories are poor candidates for a compositional semantics.

The "Mysteries of Science" Problem Not all theory-theorists claim that cognitive development mimics patterns in the history of science, but among those that do, another problem is specifying the mechanism responsible for cognitive development. Alison Gopnik and Andrew Meltzoff take up this burden by claiming that the very same mechanism is responsible for both scientific theory change and cognitive development. Yet this raises a serious difficulty: The appeal to science isn't informative if the mechanisms of theory change in science are themselves poorly understood.

Unfortunately, this is exactly the situation that we seem to be in. Gopnik and Meltzoff do their best to characterize in broad terms how one theory comes to give way to another in science. Some of their observations seem right. For instance, theories are often protected from recalcitrant data by ad hoc auxiliary hypotheses, and these eventually give way when an intense period of investigation uncovers more recalcitrant data, alongside a superior alternative theory. But how do scientists arrive at their new theories? Gopnik and Meltzoff have little more to say than that this is the "mysterious logic of discovery" (1997, p. 40). And what is distinctive about the transition from one theory to another? Here they emphasize the role of evidence and experimentation. It too is "mysterious, but that it plays a role seems plain" (p. 40). We don't doubt that experimentation is at the heart of science but without articulated accounts of how transitions between scientific theories take place, it simply doesn't help to claim that scientific and cognitive development are one and the same. Saying that two mysterious processes are really two facets of a single process is suggestive, but it hardly dispels either mystery. In other words, it's simply misleading to cite as an advantage of the Theory-Theory that it solves the problem of cognitive development when the mechanism that is supposed to do all the work is as intractable as the problem it's supposed to explain.

Like the other theories we've discussed so far, the Theory-Theory has substantial motivation and a number of serious challenges. Though it does well in explaining certain types of categorization judgments, it has trouble in allowing for stability within the conceptual system and in accounting for the referential properties of concepts. This isn't to say that there is no analogy between concepts and theoretical terms. But it does call into question whether the Theory-Theory can provide an adequate account of the nature of concepts.

Box 6

Summary of Criticisms of the Theory-Theory

1. **The Problem of Ignorance and Error**
   It is possible to have a concept in spite of its being tied up with a deficient or erroneous mental theory.
2. **The Problem of Stability**
   The content of a concept can't remain invariant across changes in its mental theory.
3. **The "Mysteries of Science" Problem**
   The mechanisms that are responsible for the emergence of new scientific theories and for the shift from one theory to another are poorly understood.

5. *The Neoclassical Theory of Concepts*

5.1. *Updating the Classical Theory*

Within psychological circles, the Classical Theory is generally considered to be a nonstarter except by those Dual Theorists who relegate classical structure to conceptual cores. In contrast, elements of the Classical Theory continue to be at the very center of discussion in other areas of cognitive science, especially linguistics and, to some extent, philosophy. We'll bring together a variety of theories emanating from these fields under the heading of the *Neoclassical Theory of Concepts*. In some ways, this family of views is the most heterogeneous in our taxonomy. Some neoclassical theorists are really just contemporary classical theorists who are sensitive to the objections we've already reviewed. Others depart from the Classical Theory on substantive points while expanding its resources in new directions. We'll say something about each of these two groups, but our focus will be on the second.

Much of the interest in the Neoclassical Theory is to be found among linguists investigating the meanings of words, especially verbs. Steven Pinker, for instance, is keenly aware that the project of specifying definitions for words is highly suspect. He notes that "[t]he suggestion that there might be a theory of verb meaning involving a small set of recurring elements might be cause for alarm" (1989, p. 167). Still, his proposal is that definitions of a sort are a perfectly viable goal for lexical semanticists (p. 168):

> I will not try to come up with a small set of primitives and relations out of which one can compose definitions capturing the totality of a verb's meaning. Rather, the verb definitions sought will be hybrid structures consisting of a scaffolding of universal, recurring, grammatically relevant meaning elements and slots for bits of [real-world knowledge]....

This view has strong affinities with the Classical Theory, in spite of its admission about real-world knowledge entering into the definition of a word. Ray Jackendoff, another neoclassical theorist, emphasizes the Classical Theory's commitment to necessary conditions but adds that a word's meaning includes other information as well (Jackendoff 1983, p. 121):

> At least three sorts of conditions are needed to adequately specify word meanings. First, we cannot do without *necessary* conditions: e.g., "red" must contain the necessary condition COLOR and "tiger" must contain at least THING. Second, we need graded conditions to designate hue in color concepts and length-width ratio of cups, for example. These conditions specify a focal or central value for a continuously variable attribute.... Third we need conditions that are typical but subject to exceptions—for instance, the element of competition in games or a tiger's stripedness.

The commitment to necessary conditions ties Jackendoff to the Classical Theory, but, like Pinker, he thinks that there are different parts to a word's meaning. This is a characteristic view among lexical semanticists, even if there is a healthy amount of disagreement about what these different parts are. Abstracting from such internal disputes, we can say that what distinguishes the Neoclassical Theory is the idea that concepts have *partial definitions* in that their structure encodes a set of necessary conditions that must be satisfied by things in their extension. Following Jackendoff, one

might hold, for example, that the structure of the concept RED embodies the condition that something can't be red without being colored. What makes this a partial definition is that this much structure encodes only a necessary condition and, at any rate, doesn't specify a sufficient condition for something's falling under the concept.

Though the appeal to partial definitions may be viewed by some as something of a cop-out, the situation isn't that lexical semanticists are just trying to put a happy face on Plato's Problem. Rather, neoclassical theorists begin with a variety of interesting linguistic phenomenon and argue that only concepts with neoclassical structure can explain this data. It may help to work through an example. Consider Jackendoff's explanation of causative constructions—a fairly standard treatment in the field of lexical semantics. Jackendoff's starting point is the observation that causatives exhibit a pronounced distributional pattern (1989 [chapter 13 in this volume], p. 50).

(16) a. x killed $y \rightarrow y$ died
b. x lifted $y \rightarrow y$ rose
c. x gave z to $y \rightarrow y$ received z
d. x persuaded y that $P \rightarrow y$ came to believe that P

Now these inferences could all be treated as having nothing to do with one another. But they are strikingly similar, and this suggests that they have a common explanation. Jackendoff's suggestion is that the meaning of a causative implicates a proprietary event and that, under this assumption, the pattern of inferences can be explained by introducing a single rule that covers all these cases, namely,

(17) X cause E to occur $\rightarrow E$ occur

For instance, the proper analysis of (16d) is supposed to be: x cause [y came to believe that P]. This analysis, taken in conjunction with the inference rule (17) implies y came to believe that P. In the present context, however, this is just to say that the concept PERSUADE has structure. CAUSE TO BELIEVE gives a partial definition of PERSUADE. There may be more to persuading someone that P than causing them to believe P,[68] but at least this provides a necessary condition for the application of PERSUADE. Moreover, this necessary condition is one that is evidenced in the distributional pattern of English illustrated by (16a)–(16d).

The causatives are just one example of how the Neoclassical Theory finds support in linguistic phenomena. Neoclassical structure has also been invoked to explain a variety of data connected with polysemy, syntactic alternations, and lexical acquisition.[69]

In philosophy, too, neoclassical structure is taken to have explanatory support. Some of the data at stake include people's intuitions about the application of a concept. Georges Rey, for example, claims that Quine's arguments against the analytic-synthetic distinction are flawed and holds, as a consequence, that it is an open question how we are to understand what he calls the *analytic data*. The analytic data

68. For example, suppose you fall down the stairs when you are walking just a bit too fast. This might lead an observer to believe that one should approach the stairs with caution. Yet, intuitively, you didn't persuade the observer of this; you merely caused him to believe it.

69. On polysemy, see Jackendoff (1989); on syntactic alternations and lexical acquisition, see Pinker (1989). For a useful collection that shows the scope of contemporary lexical semantics, see Levin and Pinker (1991b).

concern our judgments about the constitutive conditions for satisfying a concept. For example, upon hearing a Gettier example (see sec. 2), most people can be relied upon to appreciate its force; knowledge can't be (just) justified true belief. Why is it that people have this intuition? Rey's claim is that we need a theory of why this is so. "[W]e need to ask here exactly the question that Chomsky asked about syntax: what explains the patterns and projections in people's judgments?" (1993, p. 83). Rey's answer is that, by and large, the analytic intuitions are best explained by the theory that they reflect constitutive relations among our concepts. A concept such as KNOWLEDGE may have a definition after all, or at least a partial definition; it's just that the definition involves tacit rules that are extremely difficult to articulate.[70]

The Neoclassical Theory has an affinity with the Classical Theory because of its commitment to partial definitions. But the motivation for the Neoclassical Theory is largely independent of any desire to preserve the Classical Theory. The typical neoclassicist is someone who invokes partial definitions for explanatory reasons. With these motivations in mind, we turn now to some problems facing the Neoclassical Theory.

Box 7

The Neoclassical Theory

Most concepts (esp. lexical concepts) are structured mental representations that encode partial definitions, i.e., necessary conditions for their application.

5.2. *Problems for the Neoclassical Theory*

The Problem of Completers Many of the problems facing the Neoclassical Theory aren't new. In fact, it's not clear that the Neoclassical Theory offers a truly distinctive perspective on concepts at all. This comes out most vividly when we consider the question of how the partial definitions offered by neoclassical theorists are supposed to be filled out. Here neoclassical theorists confront a dilemma. On the one hand, if the partial definitions are to be turned into full definitions, then all of the problems that faced the Classical Theory return.[71] On the other hand, if they are left as partial definitions, then the Neoclassical Theory is without an account of reference determination.

We suspect that this dilemma hasn't been much of a worry among some neoclassical theorists because they aren't interested in giving a theory of concepts per se.

70. Christopher Peacocke, who in some ways is a model classical theorist (see Peacocke 1996a, 1996b [chapters 14 and 16 in this volume]), holds a similar view in a recent elaboration of his theory of concepts. See Peacocke (1997).

71. A possible exception is Katz (1997), which explicitly addresses the Problem of Analyticity. Katz argues, e.g., that the discovery that cats aren't animals is consistent with its being analytic that cats are animals. He is able to do this by claiming that, contrary to most accounts, analyticity isn't tied up with the notions of reference and truth. For Katz, analyticity is simply a matter of the containment relations among concepts. If CAT contains ANIMAL, then it's analytic that cats are animals. Whether CAT *refers* to creatures that are animals is another matter.

They are interested, instead, in grammatically relevant aspects of word meaning. For instance, when Steven Pinker claims that his "definitions" aren't intended to capture all of a verb's meaning, we take it that his point is that he isn't aiming to provide a complete characterization of the concept that the verb encodes. Understandably, given his interest in natural language, his focus is on those aspects of conceptual structure that are manifested in grammatical processes. His slots for "bits of [real-world knowledge]" are a gesture toward the larger project outside of the study of grammar, yet this is a project that Pinker is under no obligation to pursue. Jane Grimshaw is perhaps even clearer on this point. For example, she states that the words "dog" and "cat," or "melt" and "freeze," are synonymous. She doesn't mean by this that, in all senses of the term, they have the same content. The point is rather that they have the same content insofar as content has grammatical influence. "Linguistically speaking pairs like these are synonyms, because they have the same structure. The differences between them are not visible to the language" (unpublished ms., p. 2). These remarks indicate a circumscribed yet sensible research program. Grimshaw is concerned with conceptual structure, but only from the point of view of its effects on grammar. Grammatically relevant structure she calls *semantic structure*; the rest she calls *semantic content*. "Semantic structure has linguistic life, semantic content does not" (p. 2).

Still, those of us who *are* interested in the nature of concepts can't be so indifferent to the Problem of Completers. Either partial definitions are fleshed out or they are not. If they are, then the problems associated with the Classical Theory return. If they are not, then we are left without an account of how concepts apply to their instances. What makes it the case that DOG applies to all and only dogs? The fact that the concept incorporates the feature ANIMATE may place a constraint on an explanation—DOG can only apply to animates—but it is a constraint that is far too weak to answer the question.

The Problem of Ignorance and Error Because so many neoclassical theorists shy away from defending comprehensive theories of concepts, it's hard to say whether their theories are subject to the Problem of Ignorance and Error—a problem that we've seen crops up for just about everyone else. Among those neoclassical theorists who expect to complete their partial definitions, it's likely that they would have as much trouble with ignorance and error as classical theorists have. This is one respect in which the Neoclassical theory may be on the same footing as its predecessor. In both cases, there is the strong danger that the structure of a concept will encode insufficient information, or erroneous information, and so won't be able to fix the concept's reference.

Still, some neoclassical theorists may have views on reference determination that aren't readily assimilated to the Classical Theory. Ray Jackendoff's work in this area stands out. For while his theory is sensitive to grammatical indices of conceptual structure, it doesn't stop short with what Grimshaw calls semantic structure. Jackendoff's theory *is* about the nature of concepts. What's more, the structure that he takes concepts to have, in addition to their necessary conditions, isn't just a throwback to the Classical Theory. He has a number of interesting suggestions about other aspects of conceptual structure.

We won't be able to review all of his innovations, but one seems especially pertinent. Jackendoff asks the question of how to distinguish between the lexical entries

for words that are closely related in meaning, such as "duck" and "goose." He notes that these words have much the same structure in that both exhibit such general features as ANIMATE and NONHUMAN. But what makes the two have different meanings? For Jackendoff the suggestion that they differ with respect to a single additional feature is absurd; it's not as if "duck" has −LONG NECK and "goose" +LONG NECK. "To put a +/− sign and a pair of brackets around any old expression simply doesn't make it into a legitimate conceptual feature" (1989, p. 44). Jackendoff's alternative suggestion is that the lexical entries for object words include spatial information organized around a 3-D model (understood along the lines of Marr 1982). A 3-D model is a sophisticated spatial representation, but in essence, Jackendoff's theory is an elaboration of the idea that "knowing the meaning of a word that denotes a physical object involves in part knowing what such an object looks like" (Jackendoff 1987, p. 201). Though the emphasis here is on word meanings, we take it that Jackendoff's view is really about the concepts that words express. Lexical concepts for objects have a structure that incorporates a 3-D model in addition to the more mundane features that are the stock and trade of lexical semantics.

That this is Jackendoff's view of lexical concepts seems clear. On the other hand, how the view is supposed to connect with issues of reference determination is less clear. The problem is that Jackendoff has a negative attitude toward truth-theoretic semantics and generally shies away from the notion of reference. But these reservations really are beside the point. What's at stake is that a theory of concepts needs to capture a normative dimension of meaning—at a minimum, by pulling apart cases of erroneous categorization from cases of veridical categorization (see note 14). The suggestion we are entertaining is that spatial representations supplement features for necessary conditions, and that the resulting structure determines which things fall under a concept.

Unfortunately, such structure isn't up to the task, and for much the same reason that prototype structure isn't. Something can satisfy the properties specified by the spatial representation without falling under the concept, and something can fall under the concept without satisfying the properties specified by the spatial representation. For instance, an animal that strongly resembles a goose needn't be one, and a goose may for whatever reason fail to look like one. People readily appreciate this fact. Recall our 3-legged, tame, toothless, albino tigers. They are, nonetheless, recognized to be tigers. A theory of concepts that can't do justice to this fact is simply inadequate.[72]

The Regress Problem for Semantic Field Features Since the Neoclassical Theory is motivated by a diverse set of explanatory goals, its status, to a large extent, turns on how it meets the data. That is, a full evaluation of the theory would require a thorough evaluation of whether neoclassical structure is part of the best explanation of a host of linguistic phenomena. We can't provide anything of the sort here, but we will briefly discuss a methodological objection to some representative arguments in lexical semantics according to which the lexical concepts have semantic field features. These features are supposed to access patterns of inferences that are proprietary to a

72. Which isn't to say that the theory is entirely wrong. Just as prototypes might still be part of the nature of concepts even though they don't determine reference, so might 3-D representations. It's doubtful, however, that Jackendoff would want to accept a version of the Dual Theory, as many prototype theorists have.

particular field. For instance, concepts with a feature indicating the field "spatial location and motion" may license one body of inferences, while a feature indicating the field "scheduling of activities" may license another. Such differences are supposed to account for distributional patterns where lexical items that have similar meanings nonetheless permit distinct and characteristic inferences.

Ray Jackendoff, for example, argues for the existence of semantic field features on the basis of the following evidence, labeled according to four proposed fields (Jackendoff 1989, p. 37):

a. *Spatial location and motion*
   i. The bird went from the ground to the tree.
   ii. The bird is in the tree.
   iii. Harry kept the bird in the cage.
b. *Possession*
   i. The inheritance went to Philip.
   ii. The money is Philip's.
   iii. Susan kept the money.
c. *Ascription of properties*
   i. The light went/changed from green to red.
      Harry went from elated to depressed.
   ii. The light is red.
      Harry is depressed.
   iii. Sam kept the crowed happy.
d. *Scheduling of activities*
   i. The meeting was changed from Tuesday to Monday.
   ii. The meeting is on Monday.
   iii. Let's keep the trip on Saturday.

The intuition that is the basis of Jackendoff's argument is that "go," "be," and "keep" are polysemous whereby, in a given semantic field, each verb has a different though similar meaning to the one it has in any other semantic field. "The *go* sentences each express a change of some sort, and their respective terminal states are described by the corresponding *be* sentences. The *keep* sentences all denote the causation of a state that endures over a period of time. One has the sense, then, that this variety of uses is not accidental" (1989, p. 37). Jackendoff's suggestion is that these intuitions ought to be taken seriously and that the way to do this is by introducing two degrees of freedom. First, the similarities of meaning can be captured under the assumption that the similar items are associated with partially identical representations. Second, the differences in meaning can be captured under the assumption that their associated representations differ with respect to a constituent that picks out a semantic field. This constituent may then interact with inference rules that explain why a single word licenses different inferences depending on its context.

To take an example, Jackendoff's representation for the "keep" verbs all share this much structure:

(1) $[_{\text{Event}}$ CAUSE $([_{\text{Thing}}\ x], [_{\text{Event}}$ STAY $([\ ], [\ ])])]$

The way we are to understand the notation is that the word "keep" expresses a function (labeled "CAUSE") that takes two arguments (one labeled "Thing," the other

labeled "Event") onto a value (labeled "Event"), where the second argument is itself a function (labeled "STAY").[73] Semantic fields may then be indicated as subscripts on the function labels. Thus the difference between "keep" in (a-iii) and "keep" in (b-iii) is to be indicated by the subscript on "CAUSE":

(2) $[_{\text{Event}}\ \text{CAUSE}_{\text{Spatial}}\ ([_{\text{Thing}}\ x],\ [_{\text{Event}}\ \text{STAY}\ ([\ ],\ [\ ])])]$

(3) $[_{\text{Event}}\ \text{CAUSE}_{\text{Poss}}\ ([_{\text{Thing}}\ x],\ [_{\text{Event}}\ \text{STAY}\ ([\ ],\ [\ ])])]$

A full elaboration of the sentences requires filling in the variables, as we've done here for a sample sentence, (a-iii):

(4) $[_{\text{Event}}\ \text{CAUSE}_{\text{Spatial}}\ ([_{\text{Thing}}\ \text{HARRY}],\ [_{\text{Event}}\ \text{STAY}\ ([_{\text{Thing}}\ \text{THE BIRD}],\ [_{\text{Place}}\ \text{IN THE CAGE}])])]$

The thing to keep your eye on is how this notation makes explicit Jackendoff's explanation of why the different occurrences of "keep" seem both similar and different in meaning. To the extent that they are similar, this is because they share the same underlying structural template, namely, (1); to the extent that they are distinct, this is because their associated representations contain different semantic field features, as in (2) and (3).

The methodological objection that is associated with this type of explanation is one that Jerry Fodor (1998) pushes vigorously. Fodor's argument is that polysemy can't be accounted for by the interaction of a verb template and a semantic field feature because this type of explanation confronts a dilemma. Either it involves an endless regress or else the postulation of neoclassical structure is simply gratuitous. The source of the dilemma is the fact that for a verb like "keep" to retain part of its meaning across semantic fields, its semantic constituents must themselves be univocal across semantic fields. If, for example, CAUSE and THING change their meaning every time they occur in a new context, then "keep" couldn't be relied upon to retain any of its meaning. So the univocality of "keep" depends upon the univocality of, among other things, CAUSE. But are we to explain CAUSE's univocality by postulating that it too has a definition? If so, then when the same problem crops up again for its defining constituents, we'll have to postulate yet more definitions, with no end in sight. On the other hand, if CAUSE can retain its meaning across semantic fields without its having neoclassical structure, then so can "keep." "Why not say that 'keep' is univocal because it always means *keep*; just as, in order to avoid the regress, Jackendoff is required to say that 'CAUSE' is univocal because it always means *cause*" (Fodor 1998, p. 52).

This being a methodological objection, it will suffice to show that there is nothing inherently flawed in Jackendoff's strategy of argument. Whether he is right that "keep" and other verbs have neoclassical structure that implicates semantic field features is, ultimately, the question of the most interest. For present purposes, however, the primary issue is the methodological one, and on this score we see no reason why Jackendoff should be worried about Fodor's dilemma.

73. We've maintained Jackendoff's notation which may be a little confusing, since his use of capital letters resembles our use of small caps. We hope readers won't be misled into thinking that only the items designated by capitals are concepts. On the contrary, all of the items that Jackendoff's notation picks out are concepts. For example, his "Event" and "CAUSE" are both internal representations that express subpropositional contents.

As we see it, Jackendoff should take hold of the second horn. He should admit that, in principle, a word can retain aspects of its meaning across semantic fields without having neoclassical structure. That is, just as CAUSE retains its meaning, so might "keep." But just because this is the case in principle, doesn't mean that the best explanation requires that one withhold the postulation of neoclassical structure. If one has an explanatory reason to invoke neoclassical structure in some cases (but not all), then the postulation of such structure isn't the least bit gratuitous. Nor need it lead to a regress. The reason for saying that "keep" has structure needn't be applicable at all levels of representation. Maybe it simply isn't valid once one gets to the level of the concept CAUSE. In short, polysemy doesn't require neoclassical structure, but there may still be an explanatory advantage to postulating the structure. It remains for Jackendoff to demonstrate this explanatory advantage. The main point, however, is that there is no a priori reason to think that there isn't one.

In general, the merits of postulating neoclassical structure depend upon the explanations that prove the most tenable for a variety of data—not just evidence of polysemy, but also data concerning syntactic phenomena, lexical acquisition, and our intuitions about the constitutive relations among concepts.[74] We see no reason why neoclassical structure shouldn't be implicated to explain these things, but just because it is doesn't mean we've been given a full account of the nature of concepts. How partial definitions are to be filled in and how their application is to be determined remain to be seen.

Box 8

Summary of Criticisms of the Neoclassical Theory

1. **The Problem of Completers**
   If partial definitions are turned into full definitions, then the Neoclassical Theory has all the problems that are associated with the Classical Theory. If, instead, they are left incomplete, then the Neoclassical Theory has no account of reference determination.
2. **The Problem of Ignorance and Error**
   Supplementing neoclassical structure with 3-D models won't help in accounting for reference determination.
3. **The Regress Problem for Semantic Fields**
   Neoclassical structure can't explain how a word retains aspects of its meaning across different semantic fields. Either its conceptual constituents must themselves have neoclassical structure, and so on, or else no structure is needed at all.

6. *Conceptual Atomism*

6.1. *Concepts Without Structure*

All of the theories that we've covered so far disagree about the structure of concepts, but that most concepts have structure—especially lexical concepts—is an assump-

74. We've postponed the discussion of the latter until sec. 6.2, where we contrast neoclassical and atomistic accounts of the analytic data.

tion they all share. The last theory of concepts that we will discuss is unique in that it denies this assumption. As Jerry Fodor puts it (1998, p. 22; emphasis removed):

> "What is the structure of the concept DOG?" ... on the evidence available, it's reasonable to suppose that such mental representations have no structure; it's reasonable to suppose that they are atoms.

This view, which we will call *Conceptual Atomism*, is sometimes met with stark incredulity. How can lexical concepts have no structure at all? If they are atoms, wouldn't that rob them of any explanatory power? After all, in other theories, it's a concept's structure that is implicated in accounts of categorization, acquisition, and all the other phenomena that theories of concepts are usually taken to address. Defenders of Conceptual Atomism, however, are motivated by what they take to be grave failings of these other theories, especially the lack of definitions (for the Classical Theory) and the imposing difficulties of compositionality (for the Prototype Theory). In addition, conceptual atomists find support in the arguments first given by Kripke and Putnam against descriptivist theories of meaning.

As stated, Conceptual Atomism is largely a negative view. It doesn't posit concepts with classical or neoclassical structure, it doesn't posit concepts with prototype structure, and it doesn't posit concepts with theory structure. It posits concepts with no structure. This may leave one wondering what a developed version of Conceptual Atomism looks like. What's needed is a theory of how the reference of unstructured concepts is determined. For purposes of exposition, we will use Fodor's Asymmetric Dependence Theory, since it is one of the most developed in the field (see Fodor, J. A. 1990a [chapter 22 in this volume]; see also Fodor, J. A. 1990b, 1990c).

The Asymmetric Dependence Theory is a descendent of the causal-historical theories of Kripke and Putnam. The heart of the theory is the idea that the content of a primitive concept is determined by the concept's standing in an appropriate causal relation to things in the world. For Fodor, the causal relation is a nomic connection between types of concepts and the properties their tokens express. For example, the content of the concept BIRD isn't to be given by its relation to such concepts as ANIMAL, WINGS, and so on. Rather, BIRD expresses the property *bird*, in part, because there is a causal law connecting the property of being a bird with the concept BIRD.[75] This much of the theory places Fodor's account squarely in the information-based semantics tradition, according to which mental content is a species of informational content (see Dretske 1981). *Information* is basically a matter of reliable correlations. Where one type of event is a reliable cause of another, the second is said to carry information about the first. So mental content, for Fodor, requires that a concept carry information about the property it expresses. But there is more to mental content than information. As is widely recognized, there are a variety of cases where a concept is a reliable effect of things that are not in its extension. The standard case of this kind is a situation where an erroneous application of a concept is, for whatever reason, reliable. Take, for instance, a situation when viewing conditions are poor. It's a dark night, perhaps a bit foggy, and you think you see a cow in the field just beyond the road. That's to say, you apply the concept COW to the entity over there, and you do

75. The extension of the concept is then a trivial consequence of the property it expresses. Something falls under the concept BIRD just in case it instantiates the property *bird*.

so for understandable reasons—it looks like a cow. Nonetheless, it's a horse; you've misapplied your concept. That's to be expected in conditions like these, since under the conditions we are envisioning, the horse actually looks like a cow. The result is that your concept COW is the reliable effect of at least two causes: cows and horses. If, however, there is nothing more to content than information, we would not have a case of error here at all, but rather a veridical application of a concept expressing the disjunctive property *cow or horse*. In philosophical circles, this issue has come to be known as the *Disjunction Problem*.

Information-based semanticists have explored a number of ways to overcome the Disjunction Problem. Fodor's solution is to claim that certain informational relations are more basic than others and that this difference is what counts. His theory has two parts:

> (1) A concept—COW, for example—stands in a lawful relation, *L*, to the property it expresses, namely, *cow*.
> (2) Other lawful relations involving COW, L_1–L_n, are asymmetrically dependent upon the lawful relation between COW and *cow*. That is, L_1–L_n wouldn't hold but that *L* does, and not the other way around.

Thus the critical difference between the *cow*/COW law and the *horse*/COW law is that although both are reliable, the first is the more fundamental: It would obtain even if the *horse*/COW dependence did not, whereas the *horse*/COW dependence would not obtain without the *cow*/COW dependence. That's why COW expresses the property *cow* and not, as it might be, *cow or horse*.[76]

Notice that an advantage of the Asymmetric Dependence Theory is that it implies that no representation that is associated with a concept is essential to its having the content that it does. In principle, one might even have the concept COW without having the concept ANIMAL. All that is required is that there be some mechanism or other that secures the right mind-world relations. As a result, Conceptual Atomism is able to sidestep some of the most persistent difficulties that confront other theories. For instance, there needn't be a problem about ignorance and error. So long as COW is appropriately connected with *cow* (the property), it doesn't matter what you believe about cows. For much the same reason, there needn't be a problem about stability. So long as COW continues to stand in the same mind-world relation, variations in surrounding beliefs can have no effect on its content.[77]

76. We should emphasize that Conceptual Atomism shouldn't be conflated with any particular theory of reference determination and its way of dealing with the Disjunction Problem. Ruth Millikan, e.g., makes use of a theory that is similar to Fodor's but which requires certain historical facts as well. "A substance concept causally originates from the substance that it denotes. It is a concept of *A*, rather than *B*, not because the thinker will always succeed in reidentifying *A*, never confusing it with *B*, but because *A* is what the thinker has been conceptually, hence physically, tracking and picking up information about, and because the concept has been tuned to its present accuracy by causal interaction with either the members of *A*'s specific domain or with *A* itself, during the evolutionary history of the species or through the learning history of the individual" (1998 [chapter 23 in this volume], p. 63; see also Millikan 1984). For a useful overview of theories of mental content, see Crane (1995).

77. To the extent that the mind-world relation is supported by varying sets of beliefs, these can be thought of as forming an equivalence class. Each set is semantically the same as all the others since they all converge on the same mind-world relation; it's this relation, however, and not the specific belief contents, that determine a concept's content.

No doubt, these are among the chief attractions of Conceptual Atomism.[78] But, like any other theory of concepts, Conceptual Atomism isn't without its own problems. We turn to these next.

Box 9

Conceptual Atomism

Lexical concepts are primitive; they have no structure.

6.2. Problems for Conceptual Atomism

The Problem of Radical Nativism One of the most powerful motivations for developing nonatomistic accounts of concepts is a worry that is often lurking in the background, even if it is left unstated. This is the view that Conceptual Atomism involves far too strong of a commitment to innate concepts. The support for this view comes from Jerry Fodor's argument that primitive concepts have to be innate (Fodor, J. A. 1981; see also Fodor, J. A. et al. 1980). Since Conceptual Atomism says that lexical concepts are primitive, atomists would be committed to a huge stock of innate concepts, including such unlikely candidates as BROCCOLI, CARBURETOR, and GALAXY. Fodor is famous—or rather, infamous—for having endorsed this conclusion.

Now few people have been enthusiastic about embracing such a radical form of nativism, but the logic of his argument and the significance of the issue aren't to be dismissed so quickly. For example, Beth Levin and Steven Pinker speak for many people in cognitive science when they defend the need for conceptual structure (1991a, p. 4):

> Psychology ... cannot afford to do without a theory of lexical semantics. Fodor ... points out the harsh but inexorable logic. According to the computational theory of mind, the primitive (nondecomposed) mental symbols are the innate ones.... Fodor, after assessing the contemporary relevant evidence, concluded that most word meanings are not decomposable—therefore, he suggested, we must start living with the implications of this fact for the richness of the innate human conceptual repertoire, including such counterintuitive corollaries as that the concept CAR is innate. Whether or not one agrees with Fodor's assessment of the evidence, the importance of understanding the extent to which word meanings decompose cannot be denied, for such investigation provides crucial evidence about the innate stuff out of which concepts are made.

In even stronger terms, Ray Jackendoff claims to endorse the logic of Fodor's argument "unconditionally"; if a concept is unstructured, he says, it can't be learned (1989, p. 50).

78. Another is that conceptual atomists don't have to distinguish the relations among concepts that are constitutive of their content from those that merely express collateral information; for an atomist, no relations among concepts are constitutive of their content. This is one reason Fodor is such an ardent supporter of atomism. He thinks that once one admits that some relations among concepts are constitutive of their content, one is forced to admit that all are. The result is supposed to be an untenable holistic semantics (Fodor 1987; Fodor and Lepore 1992).

Let's put aside the question of whether nonatomic theories of lexical concepts are defensible. What is the reasoning behind the rest of Fodor's argument? Briefly, Fodor sees only one way that cognitive science can explain the learning of a concept. This is by postulating a mechanism whereby a new complex concept is assembled from its constituents. To take a simple example, suppose that the concept FATHER is the concept of a male parent and that the concept has the structure MALE PARENT, that is, it is literally composed of the concepts MALE and PARENT (and whatever logico-syntactic concepts may be involved). In this case, one can imagine that the acquisition of FATHER proceeds by noticing that some parents are male and by constructing a complex concept to reflect this contingency, namely, MALE PARENT (= FATHER). Notice that, in this way, the learning of FATHER takes place only on the condition that the agent previously possesses the concepts MALE and PARENT. Turning to the component concepts, MALE and PARENT, we can now ask the same question about how they are acquired. Perhaps they too decompose into simpler concepts and are acquired in much the same way as we are supposing FATHER is acquired. Yet clearly this process has to stop. Eventually decomposition comes to an end, and at that point we simply can't explain acquisition in terms of a constructive process. Since this is the only explanation of how a concept is learned, there is no explanation of how primitive concepts can be learned. Thus they must be innate.

In one form or another, this argument has led many people to be weary of Conceptual Atomism. After all, accepting the innateness of GALAXY and CARBURETOR is no small matter. Fortunately, Fodor's argument isn't sound, though not primarily for the reasons that are usually cited. What's really wrong with Fodor's position is that with his focus on conceptual structure, he fails to pose the issue of conceptual acquisition in its most fundamental terms. If to possess a concept is to possess a contentful representation, the issue of acquisition is how, given the correct theory of mental content, one can come to be in a state in which the conditions that the theory specifies obtain. To answer this question one needs to look at the acquisition process from the vantage point of a developed theory of content. One of the reasons atomistic theories may have appeared to prohibit learning is precisely because they have rarely been articulated to the point where one can ask how a mind comes to satisfy their constraints. Ironically, now that Fodor has provided a detailed atomistic theory, we can see by relation to the theory how an unstructured concept might be learned.

To explain acquisition on the Asymmetric Dependence Theory one needs an account of how the mind-world dependencies that are constitutive of content come to obtain. The key to the explanation is the notion of a *sustaining mechanism*. A sustaining mechanism is a mechanism that supports a mind-world dependency relation. For some concepts there will be sustaining mechanisms in terms of neurologically specified transducers, but the majority of concepts require sustaining mechanisms that take the form of inferential processes. The idea is that although specific inferences implicating a concept aren't constitutive of the concept's content, they nonetheless contribute to the explanation of why the concept is tokened in a variety of contexts.

Since having a concept involves having an appropriate sustaining mechanism, a psychological model of concept acquisition is to be directed at the question of how various sustaining mechanisms are acquired. Margolis (1998 [chapter 24 in this volume]) examines this question in detail and catalogs a number of distinct types of sustaining mechanisms. An interesting result of this work is that a typical sustaining mechanism for natural kind concepts implicates a *kind syndrome*—the sort of information that

one might accumulate in encountering a kind—along with a more general disposition to treat instances as members of the category only if they have the same essential property that is a reliable cause of the syndrome. The significance of this account of the sustaining mechanisms for natural kind concepts is that it readily translates into a learning model. Concept learning—at least for some natural kind concepts—proceeds by accumulating contingent, largely perceptual, information about a kind. This information, together with the more general disposition, establishes an inferential mechanism that causes the agent to token her concept under the conditions which, according to the Asymmetric Dependence Theory, are constitutive of conceptual content. Since the acquisitional process relies on a relatively general process and reflects the contingencies of experience, we think it is fair to say that this is a learning model. Such a model shows how concepts might be learned in spite of lacking semantic structure.

The exact implications of a model of this kind have yet to be worked out. Most likely, it's not one that a strict empiricist would endorse, since it seems to rely upon considerable innate machinery. At the same time, it brings Conceptual Atomism together with the idea that specific concepts needn't themselves be innate. In this way, it undermines one of the chief points of resistance to atomistic theories.[79]

The Problem of Explanatory Impotence For many theorists in cognitive science, it's close to a platitude that lexical concepts can't be primitive even if the issue of radical concept nativism is put to the side. The basis for this sentiment is the thought that Conceptual Atomism is incapable of providing illuminating accounts of psychological phenomena. Were concepts atoms, they'd lack the resources to explain anything. For instance, how can atomists make sense of categorization? Without any structure, it would seem that concepts have to be applied directly, that is, without any mediating processes. Surely this is unrealistic. But what alternatives does an atomist have?

This problem encapsulates a major challenge to Conceptual Atomism, and it is vital that atomists have a response to it. Perhaps the main thing that an atomist can say is that, for any given concept, as much structure as you like may be invoked to explain its deployment, but with one serious qualification: This structure is to be treated as being merely associated with the concept rather than constituting part of its nature.

The distinction between a representation's being merely associated with another and its being partly constitutive of the other isn't new. Just about every theory makes the same distinction, each drawing the line in its own characteristic way.[80] For instance, on the Classical Theory, a concept's constitutive structure is restricted to its relations to concepts that encode the necessary and sufficient conditions for its application. You may think that bachelors make good friends; you may even rely on this belief whenever you deploy the concept BACHELOR. But on the Classical Theory, FRIEND

79. Fodor (1998) abandons a commitment to radical concept nativism, but in a different way than we are suggesting and one that we think is ultimately inadequate. In focusing on the question of how a primitive concept can be occasioned by its instances, Fodor argues for a metaphysical view about the nature of the properties that primitive concepts express. In effect, he defines these properties relative to the effects they have on human minds. However, he says nothing about the nature of the cognitive mechanisms that are responsible for concept acquisition. That is, he doesn't say anything about how these properties have the effects on us that they do. To us, this is an unsatisfactory account, since it doesn't really address the question of how concepts are acquired. For an extended discussion of these issues, see Laurence and Margolis (ms).

80. An exception would be an extreme form of meaning holism, according to which the content of a mental representation is determined by its relation to every other representation in the cognitive system. See, e.g., Lormand (1996).

remains outside of the structure of BACHELOR simply because it's not part of the definition of BACHELOR. Like any other theorist, the atomist holds that people associate a considerable amount of information with any concept they possess. The only difference is that whereas other theorists say that much of the information is collateral (and that only a small part is constitutive of the concept itself), atomists say that *all* of it is collateral. Thus for conceptual atomists a lexical concept can be unstructured while retaining its links to the representational resources that explain how it functions.

We take it that a move like this is implicit in most discussions of Conceptual Atomism. For instance, in spite of Fodor's defense of the idea that lexical concepts are primitive, he fully acknowledges the importance of prototype structure. He writes (1981, p. 293):

> Now, what is striking about prototypes as opposed to definitions is that, whereas the evidence for the psychological reality of the latter is, as we've seen, exiguous, there is abundant evidence for the psychological reality of the former. Eleanor Rosch ... and her colleagues, in particular, have provided striking demonstrations that the prototype structure of a concept determines much of the variance in a wide variety of experimental tasks, chronometric and otherwise.... Insofar as theses get established in cognitive psychology, I think we can take the reality of prototype structures as read.

In other words, Fodor endorses the existence of prototype structure and its explanatory significance, yet he denies that this structure is part of the nature of concepts; for him it's entirely collateral.[81] For Fodor, prototypes are related to their concepts in much the way that a classical theorist would say that FRIEND is related to BACHELOR. If there is any difference, it's just that prototypes involve cognitive relations that have more reliable and pervasive effects.

The Problem of the Analytic Data As we noted earlier, one reason that philosophers cite for thinking that concepts have partial definitions is that this provides an explanation of the analytic data. People can feel the pull of a proposed definition or a counterexample and, more generally, they are able to form judgments about the constitutive conditions for satisfying a concept. George Rey (1993) has marshaled an argument against Conceptual Atomism based on this data. His claim is that quite apart from the question of whether there are any analytic truths, people certainly have intuitions about what's analytic. One explanation of these intuitions is that they reflect constitutive relations among the concepts at stake. So barring an alternative atomistic explanation, we have simultaneously an argument against Conceptual Atomism and an argument for the Neoclassical Theory. Rey's position is that no plausible atomistic alternative exists.

One atomistic proposal Rey considers is that intuitions of analyticity reflect the way that a concept is introduced. For instance, one might try to maintain that we learn a concept like BACHELOR by being told that bachelors are unmarried men. This explanation is inadequate, however, as it fails to address a range of cases where there

81. More precisely, he denies that prototypes are part of the semantic structure of concepts. Since he seems to assume that there is nothing more to the structure of a concept than its semantic structure, he doesn't distinguish between the two claims. We've seen, however, that some theorists do distinguish them (e.g., dual theorists), so one has to be careful. We'll return to the question of how to think about conceptual structure in sec. 7.

are intuitions of analyticity, and it implies that there should be intuitions of analyticity in cases where there are none. Thus, as Rey points out, few of us learned what knowledge is by being told that knowledge is (at least) justified true belief. And in spite of the fact that almost all of us had our first acquaintance with Christopher Columbus by being told that Columbus discovered America, no one has the intuition that "Columbus discovered America" is analytic.[82]

Another atomistic explanation of our intuitions of analyticity is that they merely reflect deeply held beliefs, perhaps ones that are so central to our thinking or so entrenched that we find it nearly impossible to abandon them. For instance, logical and mathematical truths have always been among the best candidates for analytic truths, and they are especially difficult to abandon. Once again, however, Rey argues that the explanation fails in both directions. On the one hand, the most compelling analyses of philosophically interesting concepts (e.g., KNOWLEDGE) are hardly entrenched; they don't even command widespread acceptance. On the other hand, many beliefs that are deeply entrenched don't seem in the least analytic (e.g., that the Earth has existed for more than five minutes).

In Rey's view it's unlikely that atomists have an adequate explanation of our intuitions of analyticity. Of course, atomists might insist that it's wrong to expect a single explanation of the intuitions. After all, from the point of view of Conceptual Atomism, the intuitions of analyticity are faulty (see, e.g., Fodor 1998). But we think there is a simpler atomistic response.

To a first approximation, the intuitions of analyticity might be explained by claiming that they reflect our entrenched beliefs about the constitutive conditions for satisfying particular concepts. That is, they don't reflect actual constitutive conditions, but rather our deeply held *beliefs* about such conditions. Notice that this theory addresses all of the cases that Rey cites. Thus we believe that it's constitutive of being a bachelor that the person be unmarried and male. But we don't believe that it's constitutive of being Columbus that he discovered America. We believe that it's constitutive of knowledge that it be at least justified true belief. But we don't believe that it's constitutive of anything that the Earth should have existed for more than five minutes.

Unfortunately, this first approximation isn't quite right. Notice that we can have entrenched beliefs about what's constitutive of what that do not seem analytic. For example, many people are totally convinced that water is H_2O—that it is constitutive of water that it has the chemical composition H_2O. Yet no one thinks it's analytic that water is H_2O. The amendment that our theory requires is that it should be intuitively or pretheoretically obvious that the condition is constitutive. That is, on our theory a belief that, say, bachelors are unmarried should seem obvious, whereas the comparable belief in the case of water/H_2O should not. And that does seem right. Even people who are thoroughly convinced that water is H_2O don't think it is obviously so; you have to know your chemistry.

So there is an atomistic alternative to the Neoclassical Theory. Moreover, our account has an advantage over the Neoclassical Theory. One of the interesting psychological facts surrounding the intuitions of analyticity is that they vary in the

82. A related suggestion, which is subject to the same counterexamples, is that intuitions of analyticity derive from a process of conditioning. That is, they aren't owing to a single introduction to a concept but to an extended process in which people are exposed to the same information, over and over again, until it's drilled in.

extent to which they hold our convictions. The examples involving BACHELOR are about as firm as they come. But other cases are less secure. Is it analytic that cats are animals? Here our own intuitions waver, and the controversies surrounding this case seem to suggest that other people's intuitions are less secure as well. Our account of the analytic data predicts this variability. Part of the variability traces back to the clause that the constitutive relation has to seem obvious; surely some things are less obvious than others. But another part traces back to the clause that the belief is entrenched. We need only add that not all such beliefs are equally entrenched. Those that are highly entrenched will give rise to firm intuitions of analyticity; those that are less entrenched will give rise to shakier intuitions. As far as we can tell, Rey has no comparable explanation. Since he relies upon actual analytic connections among concepts, they would seem to be all on a par. So at this point in the debate, Conceptual Atomism may have an advantage over the Neoclassical Theory.

The Problem of Compositionality In a sense, an atomistic theory of concepts such as Fodor's doesn't have any problem with conceptual combination. Yet this is only because, as the theory is posed, it is restricted to lexical concepts.

Suppose, however, that we treat Fodor's theory of reference determination as a comprehensive theory of concepts, in the same way that we initially treated the Prototype Theory. Then his theory appears to have difficulties that will seem all too familiar. Consider, for example, a concept we discussed in connection with the Prototype Theory, an example that's owing to Fodor himself—GRANDMOTHERS MOST OF WHOSE GRANDCHILDREN ARE MARRIED TO DENTISTS. It is hardly likely that this concept stands in a lawful dependency relation with the property of being a grandmother most of whose grandchildren are married to dentists. Nor is it likely that any other dependency relations that it might stand in are asymmetrically dependent on this one (Laurence 1993).[83]

Earlier (in sec. 3.2) we quoted Fodor and Lepore arguing against Prototype Theory in the following way:

1. Prototypes aren't compositional.
2. Concepts are compositional.
3. So concepts aren't prototypes.

But asymmetric dependence relations are in exactly the same position. The asymmetric dependence relations of complex concepts aren't a function of the asymmetric dependence relations of their constituents. Thus one could adopt an argument against the Asymmetric Dependence Theory that runs parallel to Fodor and Lepore's argument against the Prototype Theory:

1. Representations in asymmetric dependence relations aren't compositional.
2. Concepts are compositional.
3. So concepts aren't representations in asymmetric dependence relations.

Fodor, of course, is aware of the difficulties surrounding complex concepts. His own way out has two parts. The first we've already noted: He stipulates that his theory applies to lexical concepts only. The second, which is just as important, is that he appeals to a different theory to account for complex concepts. This move on his part

83. Fodor's theory also has special difficulties with any complex concept that by definition picks out items that can't be detected, e.g., UNDETECTABLE STAR BIRTH.

is crucial, since he needs some way to account for complex concepts, and asymmetric dependence won't do. The theory he ends up using is the Classical Theory. Not implausibly, Fodor claims that patently complex concepts have classical constituents.

What, then, is to stop the prototype theorist from saying the same thing? The short answer is: nothing. Prototype theorists can also stipulate that, as a theory of reference determination, the Prototype Theory only covers lexical concepts. Then once the reference for these concepts is determined, they can compose into increasingly complex concepts in accordance with classical principles.[84] Of course, the Prototype Theory may still have trouble with explaining the reference determination of lexical concepts—a problem we discussed earlier. The point here, however, is that the problems specifically associated with conceptual combination needn't be understood as giving an independent argument against the Prototype Theory. In particular, they needn't favor Conceptual Atomism over the Prototype Theory.

Finally, it is worth remarking that the Asymmetric Dependence Theory may have difficulties with a variety of concepts that have received little attention, because their interest depends to a large extent on their contributions to complex concepts. For instance, it's not the least bit clear what the Asymmetric Dependence Theory says about the semantic properties of concepts for prepositions, verbs, or adverbs. How does asymmetric dependence apply to OF or IS or QUICKLY? We can highlight the problem by briefly noting the difficulties that a comparative adjectival concept like BIG presents for the theory. Since things aren't big absolutely, but big only relative to some comparison class, it's difficult to imagine the lexical concept BIG standing in the necessary asymmetric dependence relations to determine its content. One might be tempted to suppose, instead, that it derives its semantic properties by abstraction from the complex concepts in which it figures. Perhaps concepts like BIG DOG, BIG CAT, BIG TREE, and so on stand in asymmetric dependence relations to big dogs, big cats, and so on; DOG and CAT stand in such relations as well; and the semantic properties of BIG are identified with whatever mediates between these different asymmetric dependence relations. On this account, BIG itself doesn't have its semantic properties in virtue of standing in its own asymmetric dependence relations. Its content is derived from other representations that do. Unfortunately, this solution doesn't work. The problem is that since it is not just lexical concepts that can be modified by BIG, but any concept (e.g., BIG GRANDMOTHERS MOST OF WHOSE GRANDCHILDREN ARE MARRIED TO DENTISTS), we are left with the implication that the conditions of asymmetric dependence are supposed to apply directly to an unbounded number of complex concepts—a view we have already rejected.[85]

84. As we've already noted, if a complex concept has a prototype, we will still need an explanation of why this is so. But this is a completely separate issue, one which may have nothing to do with the determination of the semantic properties of the concept.

85. A further complication—but one we'll ignore—is that, in point of fact, even concepts like CAT and DOG don't stand in simple asymmetric dependence relations with the properties they express. The problem is that concepts are tokened in the context of thoughts, and in most thought contexts a concept needn't stand in *any* lawful relations to the property it expresses. Perhaps CAT stands in a lawful relation to the property of being a cat in the context of the thought THAT'S A CAT. But it's hardly obvious that it will in contexts like CATS ARE EXTINCT OF THAT'S NOT A CAT OF even CATS ARE ANIMALS. On the contrary, it seems pretty clear that it won't in these contexts. This is actually quite a serious problem for theories of content generally, but very little has been said about it. For Fodor's attempt to address these problems, see Fodor (1990a [chapter 22 in this volume]).

The Problem of Empty and Coextensive Concepts Conceptual Atomism implies that the reference of a lexical concept isn't determined by its structure. This view contrasts with all the other theories we've looked at, in that on all the other theories, lexical concepts have structure and it's their structure that determines their reference. One way of putting the difference is that other theories of concepts are descriptivist; an item falls under a concept just in case it satisfies the description that is encoded by the concept's structure. We've seen that the advantage of a nondescriptivist theory is that it is better equipped to handle difficulties such as the Problem of Stability; but descriptivist theories have their advantages too. One is a point that will be familiar from our discussion of Frege. If all there is to the content of a concept is its reference, then there is no way to distinguish coreferential concepts. Descriptivist theories have no trouble here, since they distinguish coreferential concepts in terms of their differing structures; the structure of a concept acts as its mode of presentation. In contrast, atomic theories have considerable trouble with coreferential concepts.

To see the significance of this issue, consider a case where two concepts are coextensive as a matter of necessity. Take, for instance, the concepts TRIANGULAR and TRILATERAL. Since every geometrical object that instantiates the one must instantiate the other, it's hard to see how to pull apart the properties *triangular* and *trilateral*. Supposing that there is a law connecting *triangular* with TRIANGULAR, there must also be a law connecting *trilateral* with TRIANGULAR. But surely the latter isn't asymmetrically dependent on the former. If trilateral objects didn't cause tokenings of TRIANGULAR, how *could* triangular objects cause tokenings of TRIANGULAR?[86] To take another example, suppose, as many philosophers do, that the properties *water* and H_2O are identical. How, then, can the Asymmetric Dependence Theory distinguish between the concepts WATER and H_2O? Both would be nomically dependent upon the very same property. These considerations are all the more vivid if we consider the large stock of empty concepts that we all possess, concepts such as UNICORN and ELF. All of these concepts are correlated with the same thing, namely, nothing. Yet they are clearly distinct from one another.

Another sort of example may be of special interest to psychologists. Many species besides humans are selectively sensitive to stimuli in a way that argues that they should be credited with concepts. At the same time, it seems that the concepts they have are not always the same as our own, even when they apparently have the same extension. For instance, Richard Herrnstein and his colleagues have conducted a range of experiments where pigeons have proven to be highly skilled at sorting photographs into those that depict trees from those that do not (Herrnstein 1979, 1984). The photographs were taken from a variety of perspectives—some showing close-ups of the ends of a few branches, some showing tree-covered shores from a substantial distance, and so on. Contrasting photographs depicted close-ups of celery stalks and the like. Despite the vast differences among the photographs of trees and the existence of the tree-like items in the nontree photographs, pigeons are able to sort them with considerable accuracy. What's more, they are able to do much the same for a number of other categories, including *human, fish, flower,* and *automobile.* It looks as though they are causally responsive to groupings of objects that are very nearly coextensive with salient categories of human cognition. At the same time, it

86. Cf. also pairs of concepts such as BUY and SELL. Every event in which something is bought is also an event in which something is sold. How can Asymmetric Dependence distinguish the two?

seems unlikely that we should credit them with possessing the same concepts that we do. Does a pigeon really have the concept AUTOMOBILE?

The Asymmetric Dependence Theory does have some resources for dealing with these problems, though it doesn't have an easy time with them. Fodor (1990c) suggests that the theory can account for empty concepts like UNICORN, since laws can hold between properties even if they are uninstantiated. Though there aren't any unicorns, it may still be a law that unicorns cause UNICORN's. And laws between other types of things (e.g., horses with artificial horns) and UNICORN's may be asymmetrically dependent on the unicorn/UNICORN law.

Another suggestion of Fodor's helps with the WATER/H_2O case. Here he is willing to accept they are distinct concepts on the grounds that H_2O is actually a complex concept and, in particular, that its structure implicates the concepts HYDROGEN and OXYGEN (Fodor 1990c). So one can't have the concept H_2O without having the concept HYDROGEN, but one can have the concept WATER without having any chemical concepts. Fodor summarizes this position by saying that his theory permits that some concepts are distinguished by their inferential roles—it's just that these are ones where the complexity of the concept isn't in dispute.

Still, it remains to be seen whether the Asymmetric Dependence Theory can avoid a larger commitment to the idea that the relations among concepts are constitutive of their identity. Consider, again, the concepts TRIANGULAR and TRILATERAL. The obvious suggestion for distinguishing between them is to supplement the conditions of asymmetric dependence with a limited amount of inferential role. One could say that TRIANGULAR involves an inferential disposition that links it specifically to the concept ANGLE, whereas TRILATERAL involves a disposition that links it to the concept SIDE. Similarly, one might hold that the difference between the pigeon concepts that pick out automobiles and trees and the human concepts, AUTOMOBILE and TREE, is to be given in terms of their inferential roles. TREE and AUTOMOBILE may be tied up with other concepts (e.g., NATURAL KIND and ARTIFACT), concepts that may have no role in pigeon cognition.

We suspect that many theorists who are sympathetic to information-based semantics also want to allow that inferential roles are, to some extent, part of the nature of concepts. In a way, the suggestion is to combine the Neoclassical Theory with the theories of reference that, in the first instance, find their home among conceptual atomists.[87] From the point of view of the Neoclassical Theory, it makes perfect sense to co-opt the Asymmetric Dependence Theory, or some other information-based semantics, since as we've already seen neoclassical structure is far too limited to account for the reference of a concept. On the other hand, the sort of theory that we are imagining here departs considerably from the doctrine of Conceptual Atomism. To the extent that the relations among lexical concepts determine their identity, lexical concepts can no longer be treated as atoms. They'd have some structure, even if it's not that much.

87. In philosophy, two-factor conceptual role theories take this shape. However, not all two-factor theories develop around the same motivation. Some do emphasize the referential properties of concepts, where conceptual roles are added to solve the problems that arise with coreferential concepts (see, e.g., Rey 1996 [chapter 15 in this volume]). But others seem to emphasize conceptual roles, where a theory like Asymmetric Dependence is added only to deal with the problems that arise from so-called Twin Earth examples (see, e.g., Block 1986).

Not surprisingly, Fodor is reluctant to supplement his Asymmetric Dependence Theory with inferential roles. His alternative suggestion is that coextensive concepts can be distinguished in terms of their formal properties. Like words, concepts are objects with formal and semantic properties. So just as the words "trilateral" and "triangular" are to be distinguished by their spelling or their orthography (as well as their content), the concepts TRIANGULAR and TRILATERAL are to be distinguished by whatever properties account for their being of distinct formal types. Whether this proposal works remains to be seen. It's an interesting suggestion, however, since it pulls apart several strands in the Fregean response to coextensive concepts. In the Fregean tradition, coextensive concepts are handled by saying that they have different modes of presentations. But the notion of a mode of presentation is generally understood in terms of its relevance for semantic phenomena. Don't forget: Frege said that a mode of presentation is contained within the sense of an expression and determines its reference. Another way of looking at Fodor's treatment of coextensive concepts is that he, too, wants to say that coextensive concepts differ with respect to their modes of presentation. Fodor would only add that modes of presentation needn't be part of the content of a concept; they needn't even determine a concept's reference. They simply give us a means for dealing with Frege's puzzle. In this way, Fodor may be able to maintain the view that lexical concepts are primitive, while avoiding some of the pitfalls that go with purely referential theories of content.

This completes our survey of theories of concepts. While our discussion is by no means exhaustive, we have tried to touch on the advantages and the problems associated with the major theories of concepts that are currently under debate.[88] As we've left things, no theory stands out as providing the best comprehensive account of concepts. One reason for this may be that there are different ways for a theory of concepts to contribute to an understanding of their nature. We'll take up this question in the next section.

Box 10

Summary of Criticisms of Conceptual Atomism

1. **The Problem of Radical Nativism**
   Under Conceptual Atomism, most lexical concepts turn out to be innate, including such unlikely candidates as XYLOPHONE and CARBURETOR.
2. **The Problem of Explanatory Impotence**
   If lexical concepts are primitive, they can't explain psychological phenomena such as categorization.
3. **The Problem of the Analytic Data**
   Conceptual Atomism lacks an adequate explanation of why people have intuitions of analyticity.
4. **The Problem of Compositionality**
   Atomistic theories of concepts have as much difficulty with conceptual combination as the Prototype Theory.
5. **The Problem of Empty and Coextensive Concepts**
   If concepts are atoms and the content of a concept is just its reference, then coextensive concepts can't be distinguished. As a result, all empty concepts have the same content.

88. An important exception is the Exemplar Theory. See, e.g., the excerpt from Smith and Medin (1981 [chapter 9 in this volume]) and Estes (1994).

7. *Concluding Remarks*

To begin, consider some of the explanatory roles that have been assigned to concepts. Among other things, different theories address:

- Fast categorization
- Considered acts of categorization
- Semantic application
- The licensing of inductive inference
- Analytic inference
- Concept Acquisition
- Compositionality
- Stability

Notice that the theories we've discussed aren't equally equipped to deal with each of these. For example, the Classical Theory has trouble with categorization, especially fast categorization, even though it has a natural account of compositionality (i.e., with respect to the reference determination of complex concepts). On the other hand, the Prototype Theory does far better with fast categorization, but it has considerable trouble with compositionality. Given the diversity of these explananda—and the fact that no single theory does justice to them all—one may be tempted to abandon the hope of providing a single, comprehensive theory of concepts. We think, instead, that it would be better to step back and ask how to understand claims about the nature of concepts.

Undoubtedly, some theorists want to insist that the nature of a concept is to be given solely in terms of *compositional reference-determining structure.* On this view, the structure of a concept can consist in nothing more than its relations to those other concepts that determine its reference under a principle of semantic composition. This view is what's driving the inference from the claim that prototype structures don't compose to the claim that concepts themselves don't compose. We've seen, however, that the inference breaks down. If there is more to a concept than its prototype, then there is no reason why concepts can't compose even when their prototypes don't. In a similar vein, one of the main charges against the Classical Theory—the Problem of Typicality Effects—vanishes once it's acknowledged that not all of a concept's components need to contribute to its reference. Dual Theorists tend to suppose that a concept's identification procedure has nothing to do with reference. We might say that this other structure is *nonsemantic conceptual structure.* So we have at least two views about the nature of concepts. One is that a concept can only have structure that compositionally determines its reference. The other is that concepts can have nonsemantic structure as well.

But a commitment to nonsemantic structure raises an important question: Why think that something that purports to be part of the nonsemantic structure of a concept, like a concept's identification procedure, is in any way constitutive of its identity? Why think, for example, that the features HAS GRAY HAIR, WEARS GLASSES, etc., are constitutive of GRANDMOTHER, or that FLIES, SINGS, etc., are constitutive of BIRD? The question is motivated, in part, by the assumption that some of the information associated with a concept is irrelevant to its identity. Presumably, if people think that birds are smarter than rocks, it doesn't follow merely from this fact alone that BEING SMARTER THAN A ROCK is a feature of BIRD. What is the difference, then, between BEING SMARTER

THAN A ROCK and FLIES?[89] This challenge—to single out those relations among concepts that are constitutive of their identity—is especially difficult when one is concerned with nonsemantic components. Without the constraint that a concept's structure must contribute to its content, there may be no principled way to draw the line. One suggestion—though admittedly a sketchy one—is that a concept's structure has to be robust and theoretically significant. We aren't sure what to say in general terms about when a structure is theoretically significant. As a guideline, however, we'd suggest cases where it's universal, or nearly universal, or where its appearance is a matter of psychological necessity. To the extent that prototypes are good candidates for nonsemantic structure, this is because their deployment in fast categorization does appear to be psychologically necessary, and because particular prototypes figure in robust explanations of a variety of data. So maybe the claim that concepts have nonsemantic structure can be made to stick.

Yet another view of conceptual structure is that a concept may have components that are relevant to its semantics but not to its reference. In much this spirit, Hilary Putnam suggests that a word's meaning includes a prototype-like structure even though it plays no part in the determination of the word's reference (Putnam 1970, p. 148):

> [T]here is somehow associated with the word "tiger" a *theory*; not the actual theory we believe about tigers, which is very complex, but an oversimplified theory which describes a, so to speak, tiger *stereotype*. It describes ... a *normal member* of the natural kind. It is not necessary that we believe this theory, though in the case of "tiger" we do. But it is necessary that we be aware that *this* theory is associated with the word: if our stereotype of tiger ever changes, then the word "tiger" will have changed its meaning.

This claim easily translates into a view about concepts. The suggestion is that a concept can have structure that is partly constitutive of its content even if the structure isn't implicated in an account of the concept's reference. The thing we want to emphasize is that this is a different position than the Fregean view that there is more to the meaning of a concept than its reference. After all, it was part of the Fregean program that sense determines reference. In contrast, the present suggestion is that in addition to a reference, concepts have another aspect to their content, but one that doesn't determine their reference.[90]

Finally, a fourth way of understanding conceptual structure is in terms of the sustaining mechanisms that support a reference-determining relation, such as asymmetric dependence. On this view, one concept may be part of another's structure if the first is part of a theoretically significant sustaining mechanism associated with the second. Again, what counts as theoretically significant is a hard question. But as before, it's plausible enough to include ones that are universal (or nearly universal), or ones that appear to be a matter of psychological necessity. This might be where Jackendoff's 3-D representations find their place. Perhaps they are part of the structure of object concepts. Though they have problems determining reference, there is no reason why

89. Notice that it can't simply be a matter of distinguishing which is "psychologically real"—a suggestion that is implicit in some writings on the Dual Theory (see, e.g., Landau 1982). Both are psychologically real in that the conceptual relations have psychological effects. Surely, if you ask someone whether birds are smarter than rocks, she'd say they are.

90. In philosophy, some two-factor conceptual role theories may fall in this category.

they shouldn't be an important part of the sustaining mechanisms for many object concepts. The same goes for prototypes. (For some suggestions along these lines, see Margolis 1998.)

Box 11

Four Types of Conceptual Structure[91]

1. *Compositional Reference-Determining Structure*—structure that contributes to the content and reference of a concept via a compositional semantics.
2. *Nonsemantic Structure*—structure that doesn't contribute to the content of a concept, but does contribute significantly to some other theoretically important explanatory function of concepts.
3. *Nonreferential Semantic Structure*—structure that contributes to the content of a concept but is isolated from referential consequences.
4. *Sustaining Mechanism Structure*—structure that contributes to the content of a concept indirectly by figuring in a theoretically significant sustaining mechanism, i.e., a mechanism that supports a relation such as asymmetric dependence.

An interesting implication of these different ways of thinking about conceptual structure is that theories that appear to be in conflict may actually turn out to be good partners. We'll end by mentioning one of these possibilities, a form of the Dual Theory. The twist is that instead of using classical or theory-like cores, our suggestion is that this is the place to insert Conceptual Atomism. What allows for this arrangement is a simple refinement. In light of the varying interpretations of conceptual structure, let's say that Conceptual Atomism is the view that lexical concepts lack compositional reference-determining structure (even though they may have other types of structure and *will*, in particular, have sustaining mechanism structure).

Now different theorists have specified a number of roles for conceptual cores:

(1) Cores enter into the compositional processes that generate complex concepts.
(2) Cores determine reference.
(3) Cores act as the ultimate arbiters of categorization.
(4) Cores provide stability.[92]

Surprisingly, Conceptual Atomism does fairly well by these standards.

Compositionality We've argued that Conceptual Atomism has no difficulty with conceptual combination, since it can ultimately appeal to the Classical Theory's account. Thus, as far as compositionality goes, atomic cores and classical cores are entirely on a par.

Reference Determination While no theory offers a fully satisfactory account of reference determination, atomic theories do seem to offer an advance over

91. For each of these types of structure, there will be in principle two possible interpretations—one along the lines of the Containment Model and one along the lines of the Inferential Model (see sec. 1).

92. We've already discussed (1)–(3) in connection with Osherson and Smith (1981) and Smith et al. (1984). On stability, see Smith (1989).

descriptivist theories, including the Classical Theory and the Theory-Theory, since these face the Problem of Ignorance and Error.

Ultimate Arbiters of Categorization Atomic cores do not give a satisfactory account of our most considered judgments about category membership, so they aren't suited to be the ultimate arbiters of categorization. Arguably, however, classical cores and cores with theory structure can do no better. Given the implications of confirmation holism, it may be that nothing short of the entire belief system can act as the ultimate arbiter of categorization. At best, the Theory-Theory might allow for the claim that reflective category judgments implicate theoretical knowledge, including knowledge that implicitly involves a commitment to essentialism. And, of course, this information couldn't be part of an atomic core. But Conceptual Atomism can explain these judgments by appeal to the same theoretical beliefs, claiming they are merely associated with the concept in question or, alternatively, claiming that they are part of the nonsemantic structure of the concept, alongside its prototype. The fact that the information specified by such beliefs appears to be of great theoretical significance argues for the atomist taking the latter view.[93]

Stability Since Conceptual Atomism is not a descriptivist account, the concepts it covers are largely unaffected by changes in the beliefs that are associated with them. In contrast, the Classical Theory can't provide stability until it first overcomes the Problem of Ignorance and Error, and the Theory-Theory is notoriously poor at providing stability.

In short, atomistic cores are the best of the lot. To the extent that a version of the Dual Theory is to be preferred, it's one that brings together atomic cores with prototypes and perhaps some theory structure too, all united by a nondescriptivist account of reference.

This brings us full circle. At the beginning of our discussion, we took pains to emphasize that the study of concepts has had a rich history of interdisciplinary interaction. Also, all along we've been careful to tease apart the different explanatory goals that have accompanied the major theories. The integration of these goals yields four general ways of construing the nature of a concept. In our view, each deserves to be explored in considerable detail. No doubt, this will require further cooperation across the disciplinary boundaries of cognitive science.

8. Appendix: More on Ontology

We suspect that some philosophers may be unsatisfied with our brief discussion of the ontology of concepts, since there are other reasons than Frege's for claiming that concepts can't be mental representations. Christopher Peacocke and Georges Rey may be more representative of contemporary theorists who hold that concepts are

93. We should note that the question of whether people's knowledge in a given domain is organized around a theory is distinct from the question of whether that theory determines the content of the concepts involved. Theory-theorists usually assume that the claim about content comes for free once it's established that people have internally represented theories. But it doesn't (see Margolis 1995). For instance, one could easily maintain that an internal theory of belief subserves commonsense psychological reasoning, while also maintaining that this theory fails to determine the contents for BELIEF, DESIRE, etc. Instead their contents may be determined, for example, in accordance with an information-based semantics.

abstracta (and not mental entities). For though they are both happy to allow that mental representations have their place in the scientific study of the mind, they hold out by claiming that concepts can't be identified with mental representations. Their worry, in brief, is that mental representations and concepts exhibit too loose of a connection; so they have to be distinguished. Toward the beginning of his *A Study of Concepts*, Peacocke insists on the distinction by claiming that "It is possible for one and the same concept to receive different mental representations in different individuals" (1992, p. 3). And in a recent overview of the literature on concepts, Rey remarks in much the same spirit (1994, p. 186):

> [M]any philosophers take the view that these internal representation types would no more be identical to concepts than are the type words in a natural language. One person might express the concept CITY by the word "city," another by the word "ville"; still another perhaps by a mental image of bustling boulevards; but, for all that, they might have the same concept CITY; one could believe and another doubt that cities are healthy places to live. Moreover, different people could employ the same representation to express different concepts: one person might use an image of Paris to express PARIS, another to express FRANCE.

Notice that there are two arguments here. The first is that just as different words can express the same content (e.g., the English "cat" and the French "chat"), mental representations of different types can correspond to the same concept. This is the heart of Peacocke's position. But Rey adds a second argument, going in the other direction: A single type of mental representation might correspond to multiple concepts. That is, tokens of the same representation type might turn out to express different concepts.[94]

In our view, neither of these arguments works. Despite their initial appeal, they fail to raise any difficulties for the view that concepts are mental representations.

Take the first argument. Suppose one were to grant that different types of mental representations can express the same concept—a point to which we'll return. Still, it doesn't follow that concepts can't be identified with types of mental representations. If two or more different representations of different types express the same concept, then, of course, that concept cannot be identified with one or the other of these two types. But there is no reason why the concept shouldn't be identified with a broader, more encompassing type—one that has the mental representations of these other two types among its tokens. Just as particular Persian cats can be cats alongside Siamese cats and tabbies, so tokens of different types of representations can all be instances of a broader representation type. In short, granting that different types of internal representations can express the same concept raises no difficulties for the view that concepts are mental representations.

On the other hand, it's hardly clear that one should grant that different types of mental representations can express the same concept. Perhaps a word-like mental representation and a mental image with the same, or similar, content express different concepts. Certainly they will have substantially different inferential roles. Whether

94. For ease of exposition, we will follow Rey in using the locution that a mental representation "expresses a concept." If concepts are mental representations, however, it would be better to say that a token mental representation is an instance of a mental representation type and is a concept by virtue of instantiating that type.

these two should be treated as the same concept would seem to be an open theoretical question, not one to be settled by fiat. For instance, one would face the question of whether inferential roles are constitutive of concepts and, to the extent that they are, the question of which inferential roles are relevant to conceptual identity. Given the tremendous controversy surrounding both of these issues, it makes no sense to assume from the outset that any particular difference in inferential role is irrelevant to the issue of conceptual identity.

What about Rey's second argument, that a given type of representation might be used to express different concepts by different individuals?[95] Here too the point can be granted without abandoning the claim that concepts are mental representations. If a given type of representation, *M*, can be used to express different types of concepts, then of course we cannot identify these different concepts with *M*. But nothing stops us from identifying each of the different types of concepts (e.g., PARIS and FRANCE) with other typings of mental representations, each of which can be instantiated by instances of *M*. For example, *M* might be a representation that is typed in terms of its orthographic or imagistic properties (or some other nonsemantic property). At the same time, *M* will represent one thing or another, depending upon various other facts about it—facts about its relations to other mental representations, or perhaps facts about its causal or nomic relations to things in the world. Which concept a given instance of *M* expresses will then depend not just on its being a token of *M* but also on its typing in virtue of these other facts. In other words, concepts can still be mental representations, so long as the conditions for typing representation tokens aren't confined to a highly limited set of formal properties.

As before, though, it's hardly clear that representationalists have to be so concessive. That is, it isn't obvious that as a matter of psychological fact, a given type of representation can be used to express different concepts by different individuals. For all we know, one's image of Paris might not be suited to serve as a concept of France, even if it seems on a given occurrence that it does. Why trust introspection in such cases? Perhaps what's really going on is that one consciously entertains an image of Paris and this occasions a (distinct) mental representation of France.[96]

In short, Peacocke's and Rey's arguments don't work. We haven't been given sufficient reason to think that concepts can't be mental representations, even if we accept the assumptions they ask us to make. Granting the psychological reality of mental representations, the implications are clear: Nothing is lost by saying that concepts *are* mental representations.[97]

References

Antony, L. (1987). Naturalized Epistemology and the Study of Language. In A. Shimony and D. Nails (Eds.), *Naturalistic Epistemology* (pp. 235–257). Dordrecht: D. Reidel.

95. Or, for that matter, that a single individual might use the same type of representation to express different concepts at different times.

96. That said, it does seem likely that for at least some typings of mental representations, representations so typed should be capable of instantiating more than one concept. For example, sometimes mental representations may acquire new meanings and thereby become different concepts. But even then there is no reason to say that the concepts—old and new—aren't mental representations.

97. We would like to thank Peter Carruthers, Richard Grandy, Jean Kazez, Daniel Osherson, Sarah Sawyer, Scott Sturgeon, and Jonathan Sutton for their comments on this chapter.

Armstrong, S., Gleitman, L., and Gleitman, H. (1983). What Some Concepts Might Not Be. *Cognition*, 13, 263–308. [Chapter 10, this volume.]

Ayer, A. (1946/1952). *Language, Truth and Logic*. New York: Dover.

Baillargeon, R. (1993). The Object Concept Revisited: New Directions in the Investigation of Infants' Physical Knowledge. In C. Granrund (Ed.), *Visual Perception and Cognition in Infancy* (pp. 265–315). Hillsdale, NJ: Lawrence Erlbaum Associates. [Chapter 25, this volume.]

Barsalou, L. (1987). The Instability of Graded Structure: Implications for the Nature of Concepts. In U. Neisser (Ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization* (pp. 101–140). New York: Cambridge University Press.

Block, N. (1986). Advertisement for a Semantics for Psychology. In P. A. French, T. Uehling Jr., and H. Wettstein (Eds.), *Midwest Studies in Philosophy*, vol. 10: *Studies in the Philosophy of Mind* (pp. 615–678). Minneapolis: University of Minnesota Press.

Burge, T. (1977). Belief *De Re*. *Journal of Philosophy*, 74, 338–362.

Burge, T. (1979). Individualism and the Mental. In P. French, T. Uehling Jr., and H. Wettstein (Eds.), *Midwest Studies in Philosophy*, vol. 4: *Studies in Metaphysics* (pp. 73–121). Minneapolis: University of Minnesota Press.

Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.

Carey, S. (1991). Knowledge Acquisition: Enrichment or Conceptual Change? In S. Carey and R. Gelman (Eds.), *The Epigenesis of Mind: Essays on Biology and Cognition* (pp. 257–291). Hillsdale, NJ: Lawrence Erlbaum Associates. [Chapter 20, this volume.]

Carnap, R. (1932/1959). Uberwindung der Metaphysik durch Logische Analyse der Sprache. *Erkenntnis*, vol. 2. Reprinted as "The Elimination of Metaphysics through Logical Analysis of Language" in A. Ayer (Ed.), *Logical Positivism* (pp. 60–81). New York: The Free Press.

Chomsky, N. (1959). Review of Skinner's *Verbal Behavior*. *Language*, 35, 26–58.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Clark, E. (1973). What's in a Word? On the Child's Acquisition of Semantics in His First Language. In T. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 65–110). New York: Academic Press.

Crane, R. (1995). *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representations*. London: Penguin.

Dancy, J. (1985). *Introduction to Contemporary Epistemology*. Cambridge, MA: Blackwell.

Devitt, M. (1981). *Designation*. New York: Columbia University Press.

Devitt, M., and Sterelny, K. (1987). *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge, MA: MIT Press.

Di Sciullo, A., and Williams, E. (1987). *On the Definition of Word*. Cambridge MA: MIT Press.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Estes, W. (1994). *Classification and Cognition*. New York: Oxford University Press.

Fillmore, C. (1982). Towards a Descriptive Framework for Spatial Deixis. In R. Jarvella and W. Klein (Eds.), *Speech, Place, and Action* (pp. 31–59). London: Wiley.

Fodor, J. A. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.

Fodor, J. A. (1981). The Present Status of the Innateness Controversy. In *Representations: Philosophical Essays on the Foundations of Cognitive Science* (pp. 257–316). Cambridge, MA: MIT Press.

Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1990a). Information and Representation. In P. Hanson (Ed.), *Information, Language, and Cognition* (pp. 175–190). Vancouver: University of British Columbia Press. [Chapter 22, this volume.]

Fodor, J. A. (1990b). A Theory of Content, I: The Problem. In *A Theory of Content and Other Essays* (pp. 51–87). Cambridge, MA: MIT Press.

Fodor, J. A. (1990c). A Theory of Content, II: The Theory. In *A Theory of Content and Other Essays* (pp. 89–136). Cambridge, MA: MIT Press.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.

Fodor, J. A., Garrett, M., Walker, E., and Parkes, C. (1980). Against Definitions. *Cognition*, 8, 263–367. [Excerpted as chapter 21, this volume.]

Fodor, J. A., and Lepore, E. (1992). *Holism: A Shopper's Guide*. Cambridge, MA: Basil Blackwell.

Fodor, J. A., and Lepore, E. (1996). The Red Herring and the Pet Fish: Why Concepts Still Can't Be Prototypes. *Cognition*, 58, 253–270.

Fodor, J. D., Fodor, J. A., and Garrett, M. (1975). The Psychological Unreality of Semantic Representations. *Linguistic Inquiry*, 6, 515–532.

Foss, D. (1969). Decision Processes during Sentence Comprehension: Effects of Lexical Item Difficulty and Position upon Decision Times. *Journal of Verbal Learning and Verbal Behavior*, 8, 457–462.

Frege, G. (1892/1966). On Sense and Reference. M. Black (Tr.). In P. Geach and M. Black (Eds.), *Translations from the Philosophical Writings of Gottlob Frege* (pp. 56–78). Oxford: Blackwell.

Gelman, S., Coley, J., and Gottfried, G. (1994). Essentialist Beliefs in Children: The Acquisition of Concepts and Theories. In L. Hirschfeld and S. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 341–365). New York: Cambridge University Press.

Gelman, S., and Wellman, H. (1991). Insides and Essences: Early Understandings of the Non-Obvious. *Cognition*, 38, 213–244. [Chapter 26, this volume.]

Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, 23, 121–123.

Giaquinto, M. (1996). Non-Analytic Conceptual Knowledge. *Mind*, 105, 249–268.

Gleitman, L., Gleitman, H., Miller, C., and Ostrin, R. (1996). Similar, and Similar Concepts. *Cognition*, 58, 321–376.

Gopnik, A. (1996). The Scientist as Child. *Philosophy of Science*, 63, 485–514.

Gopnik, A., and Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Grandy, R. (1990a). Concepts, Prototypes, and Information. In E. Villanueva (Ed.), *Information, Semantics, and Epistemology*. Cambridge, MA: Blackwell.

Grandy, R. (1990b). Understanding and the Principle of Compositionality. In J. Tomberlin (Ed.), *Philosophical Perspectives*, vol. 4: *Action Theory and Philosophy of Mind* (pp. 557–572). Atascadero, CA: Ridgeview Publishing Company.

Grimshaw, J. (unpublished). Semantic Structure and Semantic Content. Rutgers University. Department of Linguistics and Center for Cognitive Science.

Hahn, H. (1933/1959). Logic, Mathematics and Knowledge of Nature. In A. Ayer (Ed.), *Logical Positivism* (pp. 147–161). New York: The Free Press.

Hampton, J. (1987). Inheritance of Attributes in Natural-Concept Conjunctions. *Memory and Cognition*, 15, 55–71.

Hampton, J. (1991). The Combination of Prototype Concepts. In P. Schwanenflugel (Ed.), *The Psychology of Word Meaning* (pp. 91–116). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Herrnstein, R. (1979). Acquisition, Generalization, and Discrimination Reversal of a Natural Concept. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 118–129.

Herrnstein, R. (1984). Objects, Categories, and Discriminative Stimuli. In H. Roitblat, T. Bever, and H. Terrace (Eds.), *Animal Cognition* (pp. 233–261). Hillsdale, NJ: Lawrence Erlbaum Associates.

Horwich, P. (1992). Chomsky versus Quine on the Analytic-Synthetic Distinction. *Proceedings of the Aristotelian Society*, 92, 95–108.

Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Jackendoff, R. (1989). What Is a Concept, That a Person May Grasp It? *Mind and Language*, 4, 68–102. Reprinted Jackendoff (1992), *Languages of the Mind: Essays on Mental Representation* (pp. 21–52). [Chapter 13, this volume.]

Jackendoff, R. (1991). The Problem of Reality. *Nôus*, 25. Reprinted Jackendoff (1992), *Languages of the Mind: Essays on Mental Representation* (pp. 157–176). Cambridge, MA: MIT Press.

Kant, I. (1787/1965). *Critique of Pure Reason*. N. Kemp Smith (Tr.). New York: St. Martin's Press.

Katz, J. (1972). *Semantic Theory*. New York: Harper and Row. [Excerpted as chapter 4, this volume.]

Katz. J. (1997). Analyticity, Necessity, and the Epistemology of Semantics. *Philosophy and Phenomenological Research*, 57, 1–28.

Keil, F. (1989). *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.

Kintsch, W. (1974). *The Representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Komatsu, L. (1992). Recent Views of Conceptual Structure. *Psychological Bulletin*, 112, 500–526.

Kripke, S. (1972/1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakoff, G. (1987). Cognitive Models and Prototype Theory. In U. Neisser (Ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization* (pp. 63–100). New York: Cambridge University Press. [Chapter 18, this volume.]

Landau, B. (1982). Will the Real Grandmother Please Stand Up? The Psychological Reality of Dual Meaning Representations. *Journal of Psycholinguistic Research*, 11(1), 47–62.

Laurence, S. (1993). *Naturalism and Language: A Study of the Nature of Linguistic Kinds and Mental Representation*. Ph.D. thesis, Rutgers University.

Laurence, S., and Margolis, E. (ms). Concepts, Content, and the Innateness Controversy.

Leslie, A. (1994). ToMM, ToBy, and Agency: Core Architecture and Domain Specificity. In L. Hirschfeld and S. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 119–148). New York: Cambridge University Press.

Levin, B., and Pinker, S. (1991a). Introduction. In *Lexical and Conceptual Semantics*. Cambridge, MA: Blackwell.

Levin, B., and Pinker, S. (1991b). *Lexical and Conceptual Semantics*. Cambridge, MA: Blackwell.

Lewis, D. (1970). How to Define Theoretical Terms. *Journal of Philosophy*, 67, 427–446.

Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, 50, 249–258.

Locke, J. (1690/1975). *An Essay Concerning Human Understanding*. New York: Oxford University Press.

Lormand, E. (1996). How to Be a Meaning Holist. *Journal of Philosophy*, 93, 51–73.

Malt, B., and Smith, E. (1984). Correlated Properties in Natural Categories. *Journal of Verbal Learning and Verbal Behavior*, 23, 250–269.

Margolis, E. (1994). A Reassessment of the Shift from the Classical Theory of Concepts to Prototype Theory. *Cognition*, 51, 73–89.

Margolis, E. (1995). The Significance of the Theory Analogy in the Psychological Study of Concepts. *Mind and Language*, 10, 45–71.

Margolis, E. (1998). How to Acquire a Concept. *Mind and Language*, 13, 347–369. [Chapter 24, this volume.]

Margolis, E., and Laurence, S. (ms). Concepts as Mental Representations.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman and Company.

Medin, D. (1989). Concepts and Conceptual Structure. *American Psychologist*, 44, 1469–1481.

Medin, D., Goldstone, R., and Gentner, D. (1993). Respects for Similarity. *Psychological Review*, 100, 254–278.

Medin, D., and Ortony, A. (1989). Psychological Essentialism. In S. Vosniadou and A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 179–195). New York: Cambridge University Press.

Medin, D., and Shoben, E. (1988). Context and Structure in Conceptual Combination. *Cognitive Psychology*, 20, 158–190.

Mervis, C., Catlin, J., and Rosch, E. (1976). Relationships among Goodness-of-Example, Category Norms, and Word Frequency. *Bulletin of the Psychonomic Society*, 7, 283–284.

Miller, G. (1978). Semantic Relations among Words. In M. Halle, J. Bresnan, and G. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 60–118). Cambridge, MA: MIT Press.

Miller, G., and Johnson-Laird, P. (1976). *Language and Perception*. Cambridge, MA: Harvard University Press.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Millikan, R. (1998). 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. [Chapter 23, this volume.]

Murphy, G., and Medin, D. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review*, 92(3), 289–316. [Chapter 19, this volume.]

Osherson, D., and Smith, E. (1981). On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition*, 9, 35–58. [Chapter 11, this volume.]

Osherson, D., and Smith, E. (1982). Gradedness and Conceptual Combination. *Cognition*, 12, 299–318.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.

Peacocke, C. (1996a). Précis of *A Study of Concepts*. *Philosophy and Phenomenological Research*, 56, 407–411. [Chapter 14, this volume.]

Peacocke, C. (1996b). Can Possession Conditions Individuate Concepts? *Philosophy and Phenomenological Research*, 56, 433–460. [Excerpted as chapter 16, this volume.]

Peacocke, C. (1997). Implicit Conceptions, Understanding and Rationality. Sociedad Filosofica Ibero Americana (SOFIA), 10th Annual Conference, June 1997. Barcelona, Spain.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.

Plato (1981). Euthyphro. In G. Grube (Ed. and tr.), *Five Dialogues* (pp. 5–22). Indianapolis: Hackett Publishing Co. [Chapter 2, this volume.]

Putnam, H. (1962). The Analytic and the Synthetic. In H. Feigl and G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Science*, vol. 3. Minneapolis: University of Minnesota Press.

Putnam, H. (1970). Is Semantics Possible? In H. Kiefer and M. Munitz (Eds.), *Language, Belief and Metaphysics* (pp. 50–63). New York: State University of New York Press. [Chapter 7, this volume.]

Putnam, H. (1975). The Meaning of Meaning. In K. Gunderson (Ed.), *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.

Quine, W. (1935/1976). Truth by Convention. In *The Ways of Paradox and Other Essays* (pp. 77–106). Cambridge, MA: Harvard University Press.

Quine, W. (1951/1980). Two Dogmas of Empiricism. In *From a Logical Point of View: Nine Logico-Philosophical Essays* (pp. 20–46). Cambridge, MA: Harvard University Press. [Chapter 5, this volume.]

Quine, W. (1954/1976). Carnap and Logical Truth. In *The Ways of Paradox and Other Essays* (pp. 107–132). Cambridge, MA: Harvard University Press.

Ramsey, F. (1929/1990). Theories. In D. H. Mellor (Ed.), *Philosophical Papers* (pp. 112–136). New York: Cambridge University Press.

Rey, G. (1983). Concepts and Stereotypes. *Cognition*, 15, 237–262. [Chapter 12, this volume.]

Rey, G. (1993). The Unavailability of What We Mean: A Reply to Quine, Fodor, and Lepore. In J. A. Fodor and E. Lepore (Eds.), *Holism: A Consumer Update* (pp. 61–101). Atlanta: Rodopi B. V.

Rey, G. (1994). Concepts. In S. Guttenplan (Ed.), *A Companion to the Philosophy of Mind* (pp. 185–193). Cambridge, MA: Blackwell.

Rey, G. (1996). Resisting Primitive Compulsions. *Philosophy and Phenomenological Research*, 56, 419–424. [Chapter 15, this volume.]

Rips, L. (1995). The Current Status of Research on Concept Combination. *Mind and Language*, 10, 72–104.

Rosch, E. (1973). On the Internal Structure of Perceptual and Semantic Categories. In T. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 111–144). New York: Academic Press.

Rosch, E. (1978). Principles of Categorization. In E. Rosch and B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates. [Chapter 8, this volume.]

Rosch, E., and Mervis, C. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7, 573–605.

Rosch, E., Mervis, C., Gray, W., Johnson, D., and Boyes-Braem, P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology*, 8, 382–439.

Sellars, W. (1956). Empiricism and the Philosophy of Mind. In H. Feigl and M. Scriven (Eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis: Minnesota Studies in the Philosophy of Science* (pp. 253–329). Minneapolis: University of Minnesota Press.

Shepard, R. (1974). Representation of Structure in Similarity Data: Problems and Prospects. *Psychometrika*, 39, 373–421.

Smith, E. (1989). Concepts and Induction. In. M. Posner (Ed.), *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Smith, E. (1995). Concepts and Categorization. In E. Smith and D. Osherson (Eds.), *Thinking: An Invitation to Cognitive Science*, Vol. 3, second edition (pp. 3–33). Cambridge, MA: MIT Press.

Smith, E., and Medin, D. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Smith, E., Medin, D., and Rips, L. (1984). A Psychological Approach to Concepts: Comments on Rey's "Concepts and Stereotypes." *Cognition*, 17, 265–274.

Smith, E., Osherson, D,. Rips, L., and Keane, M. (1988). Combining Prototypes: A Selective Modification Model. *Cognitive Science*, 12, 485–527. [Chapter 17, this volume.]

Smith, E., Shoben, E., and Rips, L. (1974). Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions. *Psychological Review*, 81(3), 214–241.

Spelke, E. (1990). Principles of Object Perception. *Cognitive Science*, 14, 29–56.

Stich, S. (1993). Moral Philosophy and Mental Representation. In M. Hechter, L. Nadel, and R. Michod (Eds.), *The Origin of Values* (pp. 215–228). Hawthorne, NY: Aldine de Gruyter.

Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84(4), 327–352.

Wellman, H. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

Wittgenstein, L. (1953/1958). *Philosophical Investigations*. 3d edition. Anscombe (Tr.). Oxford: Blackwell. [Excerpted as chapter 6, this volume.]

Zadeh, L. (1965). Fuzzy Sets. *Information and Control*, 8, 338–353.