The Blackwell Guide to Philosophy of Mind
Edited by Stephen P. Stich, Ted A. Warfield

Chapter 8

Concepts

Eric Margolis and Stephen Laurence

The human mind has a prodigious capacity for representation. We aren't limited to thinking about the here and now, just as we aren't limited to thinking about the objects and properties that are relevant to our most immediate needs. Instead, we can think about things that are far away in space or time (e.g., Abraham Lincoln, Alpha Centauri) and things that involve considerable abstraction from immediate sensory experience (e.g., democracy, the number pi). We can even think about things that never have or never will exist in the actual world (e.g., Santa Claus, unicorns, and phlogiston). One of the central questions in the history of philosophy has been how we are able to do this. How is it that we are able to represent the world to ourselves in thought? In answering this question, philosophers and psychologists often take our capacity for thought to be grounded in our conceptual abilities. Thoughts are seen as having constituents or parts, namely, concepts.[1] As a result, all of science, literature, and the arts – as well as everyday thought – can be seen to stem from the astounding expressive power of the human conceptual system.

Given the foundational role that concepts have for understanding the nature of cognition, it's not possible to provide a theory of concepts without taking sides on a number of fundamental questions about the mind. In fact, the theory of concepts has become a focal point for demarcating vastly different approaches to the mind and even different worldviews. For example, it interacts with such questions as whether there really are thoughts at all and whether semantic properties are relevant to the study of human action.[2] Similarly, it is at the root of the disagreement about whether philosophy is an a priori enterprise. Needless to say, we will not discuss all of these sorts of issues here. In order to keep the discussion focused and manageable it will be necessary to make certain assumptions about matters that remain controversial both within the philosophy of mind in general and within the theory of concepts in particular.[3]

The theory of concepts has been one of the most active areas of research in both philosophy and psychology in the past 50 years, with many important and

lasting results. In what follows, we will survey a number of the most influential theories with an eye toward the key issue that divides them – the issue of conceptual structure.[4] We will argue that none of the various types of conceptual structure currently on offer is entirely satisfactory. This has led us to rethink the nature of conceptual structure itself and to distinguish several categorically different types of structure.

8.1 Definitional Structure

Theorizing about the nature of concepts has been dominated since antiquity by an account known as the *Classical Theory* of concepts. So dominant has this account been that it was not until the 1970s that serious alternatives first began to be developed. Moreover, though these alternative theories are in some respects radically different from the Classical account, they are all deeply indebted to it. In fact, it would hardly be an exaggeration to say that all existing theories of concepts are, in effect, reactions to the Classical Theory and its failings. So appreciating the motivations for the Classical Theory and its pitfalls is essential to understanding work on the nature of concepts.

According to the Classical Theory, concepts are complex mental representations whose structure generally[5] encodes a specification of necessary and sufficient conditions for their own application.[6] Consider, for example, the concept BACHELOR. The idea is that BACHELOR is actually a complex mental representation whose constituents are UNMARRIED and MAN. Something falls under, or is in the extension of, BACHELOR just in case it satisfies each of these constituent concepts. Or, to take another example, the concept KNOWLEDGE might be analyzed as JUSTIFIED TRUE BELIEF. In that case, something falls under the concept KNOWLEDGE just in case it is an instance of a true belief that's justified.[7]

This simple and intuitively appealing theory has much to recommend it. A good deal of the power and elegance of the theory derives from the fact that it is able to provide accounts of a variety of key psychological phenomena, accounts that seamlessly mesh with the treatment of reference determination just sketched. Categorization, for example, is one of the most fundamental of all processes involving concepts. Most of our higher cognitive abilities – not to mention our own survival – depend upon our ability to quickly and reliably determine which categories different objects in our environment belong to. The Classical Theory's account of this capacity is natural and compelling. What happens in categorizing something as a bird, for example, is that one accesses and decomposes the concept BIRD and checks whether its constituents apply to the object in question. If each does, then the object is deemed a bird; if at least one doesn't, then the object is not. The Classical Theory offers an equally powerful account of concept learning. The process of concept learning works in much the same way as categorization, but the process runs backwards. That is, to acquire a concept one starts

out with its constituents and assembles them in light of one's experience. Learning, on this view, is a constructive operation. One has certain concepts to begin with and brings these together to form novel, complex concepts. In short, the Classical Theory offers an elegantly unified account of reference determination, categorization, and learning.[8]

As attractive as it may be, the Classical Theory has few adherents today. This is because it faces a number of extremely challenging objections. In the remainder of this section we briefly review some of these objections to bring out certain motivations behind competing theories and to highlight a number of themes that will be relevant later on.

Perhaps the most pressing objection to the Classical Theory is the sheer lack of uncontroversial examples of definitions. This wouldn't be such a problem if the Classical Theory were part of a new research program. But the truth is that in spite of more than two thousand years of intensive sustained philosophical analysis, there are few, if any, viable cases where a concept can be said to have been defined. In fact, the failures of this research program are notorious.

To take one well-known example, consider the definition that we cited a moment ago for the concept KNOWLEDGE. The proposal was that KNOWLEDGE can be analyzed as JUSTIFIED TRUE BELIEF. As plausible as this definition sounds at first, it is subject to a family of powerful counterexamples, first noticed by Edmund Gettier. The following example is adapted from Dancy (1985). Henry is following the Wimbledon men's singles tournament. He turns on the television to watch the final match and sees McEnroe triumph over Connors. As a result, Henry comes to believe that McEnroe won the match and he has every reason to infer that McEnroe is this year's champion. But what Henry doesn't know is that, due to a problem with the network's cameras, the game can't be shown as it takes place and, instead, a recording of last year's game is being shown. Still, at this year's tournament, McEnroe repeats last year's performance, beating Connors in the final match. So Henry's belief that McEnroe is this year's champion is true and justified as well, but few people would want to say that he *knows* that McEnroe is champion this year.

It's not just philosophically interesting concepts that have problems like this. As Wittgenstein famously argued in his *Philosophical Investigations*, ordinary concepts don't seem to be any more definable than philosophical ones. One of Wittgenstein's main examples is the concept GAME, for which he considers a number of initially plausible definitions, each of which ends up being subject to a devastating counterexample. Even philosophy's stock example, BACHELOR, isn't unproblematic. Is the Pope a bachelor? How about a self-declared gay man who lives with his lover in a monogomous long-term relationship? Both are cases of unmarried men, yet neither seems to be a bachelor.

Defenders of the Classical Theory could respond that while definitions are indeed hard to come by, this doesn't necessarily mean that there aren't any. Perhaps definitions are tacit and so not easily accessible to introspection (see, e.g., Rey 1993; Peacocke 1998). The general feeling, however, is that the most likely reason why definitions are so hard to find is simply that there aren't any.

Another problem for the Classical Theory is that, because of its commitment to definitions, it is also committed to a form of the analytic/synthetic distinction – a distinction which, in the wake of Quine's famous critique, is thought by many philosophers to be deeply problematic. One strand of Quine's criticism centers around his view that confirmation is holistic. Confirmation involves global properties such as simplicity, conservatism, overall coherence, and the like. Moreover, since confirmation relies upon auxiliary hypotheses, when a theoretical claim is confronted by recalcitrant data, one can't say in advance whether it's this claim rather than some auxiliary hypothesis that needs to be abandoned. All of this seems to show that we don't have a priori access to truths that are within the realm of scientific investigation. Moreover, we don't know in advance just how far the reach of science is. What may look like a conceptual necessity (and therefore look analytic and immune to revision) may turn out to be a case where people are being misled by their own lack of theoretical imagination.

Notice, however, that if a concept has a definition, this definition will strongly constrain theoretical developments in science and place a priori limits on what we are capable of discovering about the world. For example, if the proper analysis of STRAIGHT LINE were SHORTEST DISTANCE BETWEEN TWO POINTS, then, it would seem, one couldn't discover that a straight line isn't always the shortest distance between two points. And if the proper analysis of CAT were (SUCH AND SUCH TYPE OF) ANIMAL, then one couldn't discover that cats aren't animals. These sorts of definitions would seem to be about as plausible and unassailable as they come. Yet, as Hilary Putnam (1962) has pointed out, the situation isn't so simple. With the discovery that space is non-Euclidian, we can now see that the first definition is actually wrong. And with the help of a little science fiction, we can see that it at least seems possible to discover that the second is wrong too. (Perhaps cats are actually Martian-controlled robots, and not animals at all.) But if STRAIGHT LINE and CAT had the definitions that the Classical Theory suggests, then these discoveries would be entirely prohibited; they wouldn't be possible at all. Examples like these threaten the very foundations of the Classical Theory. A definition may appear to capture the structure of a concept, but the appearance may only be an illusion which later discoveries help us to see beyond.[9]

Related to cases such as these, one finds other considerations that argue against definitions – in particular, Saul Kripke's and Hilary Putnam's influential work on the semantics of names and natural kind terms (see esp. Kripke 1972/1980; Putnam 1970, 1975). Kripke's and Putnam's target was the description theory of reference, according to which someone is able to use a name or kind term by virtue of knowing a description that picks out its reference. Notice, however, that the Classical Theory just is a form of the description theory, only it holds at the level of concept not words. For this reason, all of Kripke's and Putnam's arguments are pertinent to its evaluation. One of their arguments is an elaboration of the Quinean point that we can make discoveries about a kind that reveal that we were wrong about its nature – the problem of error. Closely related is the problem of ignorance: if people are sometimes wrong about certain properties of a

kind, they are also often ignorant of the features that really are essential to it.[10] What turns out to be crucial to the identity of gold is its atomic number, and not, for example, its color, or weight. Similarly, the crucial feature of the bubonic plague is its bacterial source, and not the chills, fever, or nausea that it is associated with, and certainly not a connection with sinful deeds (in spite of the widespread belief that the plague was a form of divine retribution). What bears emphasizing here is that such ignorance doesn't prevent people from possessing the concept GOLD or PLAGUE. If it did, people wouldn't be able genuinely to disagree with one another about the cause of the plague; they'd always end up talking at cross purposes.

The philosophical considerations weighing against the Classical Theory are impressive. But its worries don't end there. The Classical Theory also faces a number of daunting problems based on psychological considerations.

Perhaps the most glaring of these is that definitions have failed to show up in experimental situations that are explicitly designed to test for the psychological complexity of concepts (see, e.g., Kintsch 1974; J. D. Fodor et al. 1975; J. A. Fodor et al. 1980). If, for example, CONVINCE is analyzed as CAUSE TO BELIEVE (following standard Classical treatments), one would expect that CONVINCE would impose a greater processing burden than BELIEVE; after all, CONVINCE is supposed to have BELIEVE as a constituent. Yet this sort of effect has never been demonstrated in the laboratory. Not only do definitions fail to reveal themselves in processing studies, there is also no evidence of them in lexical acquisition either (Carey 1982). Of course it is always possible that these experiments aren't subtle enough or that there is some other explanation of why definitions fail to have detectable psychological effects. But it certainly doesn't help the Classical Theory's case that definitions refuse to reveal themselves experimentally.

The most powerful psychological arguments against the Classical Theory, however, are based upon so-called *typicality effects*. Typicality effects are a variety of psychological phenomena connected to the fact that people willingly rate subcategories for how typical or representative they are for a given category. For example, subjects tend to say that robins are better examples of the category *bird* than chickens are; i.e., they say robins are more "typical" of *bird*. In and of itself, this result may not be terribly interesting. What makes typicality judgments important is the fact that they track a variety of other significant psychological variables (for reviews, see Rosch 1978; Smith and Medin 1981; for a more critical review, see Barsalou 1987).

Eleanor Rosch and Carolyn Mervis (1975) found that when subjects are asked to list properties that are associated with a given category and its subordinates, the distribution of properties on these lists is predicted by independent typicality rankings. The more typical a subordinate is judged to be, the more properties it will share with other exemplars of the same category. For instance, robins are taken to have many of the same properties as other birds, and, correspondingly, robins are judged to be highly typical birds; in contrast, chickens are taken to have fewer properties in common with other birds, and chickens are judged to be

less typical birds. Another finding is that typicality has a direct reflection in categorization. In cases where subjects are asked to judge whether an X is a Y, independent measures of typicality predict the speed of correct affirmatives. Subjects are quicker in their correct response to "Is a robin a bird?" than to "Is a chicken a bird?" Error rates, as well, are predicted by typicality. The more typical the probe (X) relative to the target category (Y), the fewer the errors. Typicality also correlates with lexical acquisition and a variety of other phenomena, such as the order in which subjects will provide exemplars for a given category – more typical items are cited first. In sum, typicality effects seem to permeate every aspect of a concept's life, significantly determining its acquisition, use, and even misuse. It's no wonder that psychologists have required that a theory of concepts do justice to these data.

It's in this context that most psychologists have given up on the Classical Theory. The problem is that the Classical Theory simply has nothing to say about any of these phenomena. The classical models of categorization and concept acquisition that we sketched above don't predict any of the effects, and classical attempts to accommodate them appear ad hoc and quickly run into further problems. Moreover, as we'll see in the next section, there are alternative theories of concepts that provide natural and highly explanatory accounts of the full range of typicality effects.

The Classical Theory faces a battery of powerful philosophical and psychological objections. Definitions are very hard to come by, they don't have any psychological effects, they can't explain any of the most significant psychological facts that are known about concepts, they fly in the face of Quine's critique of the analytic–synthetic distinction, and they aren't equipped to explain how the reference of a concept is determined. As a result, it's hard to resist the thought that, in spite of its considerable attractions, the Classical Theory isn't worth saving.

8.2 Probabilistic Structure

The 1970s saw the development of a new theory of concepts, one that gained considerable support as an alternative to the Classical Theory. This new theory – the *Prototype Theory* – gave up on the idea that a concept's internal structure provides a definition of the concept.[11] Instead, the Prototype Theory adopted a probabilistic treatment of conceptual structure. According to the Prototype Theory, most lexical concepts are complex mental representations whose structure encodes not defining necessary and sufficient conditions, but, rather, conditions that items in their extension *tend to have*. So in contrast with the Classical Theory, for an object to be in the extension of a concept, it needn't satisfy each and every property encoded in the concept's structure as long as it satisfies a sufficient number of them.

Notice, right off, that one of the advantages of the Prototype Theory is that it doesn't require that concepts have definitions. It's no problem for the Prototype

Theory that people have had so much difficulty formulating them. According to the Prototype Theory, concepts, by and large, lack definitional structure; they have prototype structure instead. For this reason, it also shouldn't be a surprise that definitions never show up in studies of psychological processing. In fact, it's when we turn to the empirical psychological data that Prototype Theory becomes especially appealing. The way the theory is generally understood, it takes categorization to be a feature-matching process where an exemplar or individual is compared to a target category for how similar they are. So long as enough features match, they are deemed sufficiently similar and one comes to judge that the item falls under the category. This reliance on similarity provides the resources for an extremely natural explanation of the typicality phenomena (see, e.g., Smith 1995). One need only assume that typicality judgments are also formed by the very same process. In other words, the reason why robins are judged to be more typical birds than chickens is because ROBIN shares more features with BIRD; it ranks higher in the similarity-comparison process.

Consider also the finding by Rosch and Mervis, that typicality judgments track the number of features that a concept shares with other exemplars for a superordinate category. Again, the Prototype Theory has a natural explanation of why this happens. The reason is because the properties that subjects list that are common among the subordinate categories correspond to the features of the superordinate concept; that is, they characterize the structure of the superordinate concept. As a result, concepts that share many features with their fellow subordinates will automatically share many features with the superordinate. Sticking to the example of the concept BIRD, the idea is that the properties that are commonly cited across categories such as *robin*, *sparrow*, *ostrich*, *hawk*, and so on, are the very properties that are encoded by the structure of BIRD. Since ROBIN has many of the same structural elements, and CHICKEN has few, robins will be judged to be more typical birds than chickens are.

In short, the Prototype Theory has tremendous psychological advantages. It's no wonder that the psychological community embraced the theory as an alternative to the Classical Theory. But the Prototype Theory isn't without its difficulties either, and a full appreciation of some of these difficulties is essential to arriving at a satisfactory theory of concepts. To keep things brief, we'll mention only three.

The first problem is that the Prototype Theory is subject to the problems of ignorance and error, just like the Classical Theory. Once again, the problem is that people can possess a concept and yet have erroneous information about the items in its extension or lack a sufficient amount of correct information to pick them out uniquely. Moreover, prototypes are notoriously bad in dealing with the question of reference determination. Take, for example, the concept GRANDMOTHER. Prototypical grandmothers are women with gray hair, they have wrinkled skin, they wear glasses, and so on. Yet we all know that there are people who fail to exhibit these characteristics who are grandmothers, and that there are people who do exhibit these characteristics who are not. Mrs. Doubtfire (the Robin Williams character) may look like a grandmother, but Tina Turner really is a grandmother.

The second problem is that many concepts simply lack prototypes. This is especially clear in the case of certain complex concepts. As Jerry Fodor puts it: "[T]here 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 grandmothers*, and there are no prototypical properties of *grandmothers most of whose grandchildren are married to dentists*" (1981: 297; see also Fodor 1998).

The third problem is that prototypes don't appear to compose in accordance with the principles of a compositional semantics (see Fodor 1998; Fodor and Lepore 1996). The difficulty is that, on the standard account of how the conceptual system is productive (i.e., of how we are capable of entertaining an unbounded number of concepts), concepts must have a compositional semantics. Fodor illustrates the argument with the concept PET FISH. The PET prototype encodes properties that are associated with dogs and cats, and the FISH prototype encodes properties that are associated with things like trout, yet the PET FISH prototype encodes properties that are associated with goldfish and other small colorful fish. So it's hard to see how the prototype for PET FISH could be computed from the prototypes for PET and FISH.

Together, these three criticisms pose a serious threat to the Prototype Theory. However, prototype theorists do still have some room to maneuver. What all three objections presuppose is that prototype theorists must hold that a concept's structure is *exhausted* by its prototype. But prototype theorists could simply abandon this constraint. They could maintain, instead, that a concept's prototype is a crucial part of its structure, but that there is more to a concept than its prototype.

In fact, a number of prototype theorists have suggested theories along just these lines in order to deal with the first of our three criticisms, viz., the problem that prototypes aren't suited to determining reference. According to this *Dual Theory*, a concept has two types of structure, one type constitutes the concept's "core" and the second its "identification procedure" (Osherson and Smith 1981; Smith et al. 1984; Landau 1982). Prototypes are supposed to be confined to identification procedures. They account for quick categorization processes as well as all of the typicality effects. On the other hand, cores are supposed to have some other type of structure that accounts for reference determination and is responsible for our most considered categorization judgments – the default view being that cores exhibit classical structure.[12]

The Dual Theory handles the first objection by its commitment to conceptual cores. The idea is that it's perfectly fine if prototypes can't determine reference, since by hypothesis cores fulfil that role. It handles the second objection by adding that some concepts lack prototypes but that this doesn't prohibit anyone from possessing the concepts; they need only grasp the cores of these concepts. Finally, it handles the third objection by maintaining that the productivity of the conceptual system is established so long as conceptual cores combine in accordance with a compositional semantics, and that examples such as PET FISH don't tell against this possibility.

Though none of these responses is without merit, notice that they work by insulating prototype structure from many of the theoretical roles for which conceptual structure is introduced in the first place. As a result, the Dual Theory places a great deal of weight on the conceptual structure associated with a concept's core. To the extent that this other structure is supposed to be *classical* structure, the Dual Theory inherits most of the problems that were associated with the Classical Theory. For example, the Dual Theory faces the problem of ignorance and error, it has to overcome Quinean objections to the analytic–synthetic distinction, it has to confront the difficulty that there are few examples of true definitions, and so on. In short, the Dual Theory may expand the logical space somewhat, but, without an adequate account of conceptual cores, it isn't much of an improvement on either the Classical Theory or the Prototype Theory.

8.3 Theory Structure

The Dual Theory continues to enjoy widespread support in spite of these difficulties. We suspect that this is because of the feeling that psychology has found a way to abandon its residual ties to the Classical Theory. The idea is that conceptual cores should be understood in terms of the *Theory Theory* (see, e.g., Keil 1994). This is the view that concepts are embedded in mental structures that are in important ways like scientific theories and that they apply to the things that satisfy the descriptive content given by the roles that they have within their respective mental theories (see, e.g., Carey 1985; Murphy and Medin 1985; Gopnik and Meltzoff 1997).[13] For a mental structure to be theory-like, it must embody an explanatory schema, that is, a set of principles or rules that a thinker uses in trying to make sense of an event in the course of categorizing it. Examples of such theories include so-called common-sense psychology, common-sense physics, and common-sense biology – the sets of principles that ordinary people use in explaining psychological, physical, and biological events.[14]

One of the main advantages of the Theory Theory is the model 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 Medin and 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 (including, perhaps, relational and historical properties). The Theory Theory readily accommodates psychological essentialism since the Theory Theory takes people to appeal to a mentally represented theory in making certain category decisions. Rather than passing quickly over a check-list of properties, people ask whether the item has the right hidden property. 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: 184). This suggests that different people represent different sorts of information in thinking 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 and know very little about genetics in general.

The Theory Theory is best suited to explaining our considered acts of categorization. What matters in such cases is not so much an object's gross perceptual properties, but, rather, the properties that are taken to be essential to its nature. At the same time, the Theory Theory is not terribly well suited to explaining our more rapid categorization judgments where concepts are deployed under pressures of time and resources. And in general, the Theory Theory makes little contact with typicality effects; like the Classical Theory, it has nothing to say about why some exemplars seem more typical than others and why typicality correlates with so many other variables. On the other hand, if the Theory Theory were combined with Prototype Theory, the resulting version of the Dual Theory would seem to have considerable promise. Cores with theory structure would seem to be a vast improvement on cores with classical structure.

Unfortunately, this revised Dual Theory still faces a number of serious difficulties. We will mention two that are specifically associated with the Theory Theory as an account of conceptual cores. The first problem is one that has already cropped up, so it shouldn't be much of a surprise (the problem of reference determination); the other problem is new (the problem of stability).

The problem of reference determination affects the Theory Theory in several ways. For one thing, 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 to the extent that this is true, concepts will most likely encode inadequate information to pick out a correct and determinate extension. If people don't represent an essence for cats or dogs apart from some thin ideas about genetic endowment, then the concepts CAT and DOG will be embedded in theories that look about the same. Depending on how anemic the theories are, there may then be nothing to pull apart their concepts CAT and DOG.

On the other hand, people may have detailed enough theories to differentiate any number of concepts, yet this comes with the danger that they may have incorporated incorrect information into their theories. To return to our earlier example, someone might hold that the plague is caused by divine retribution, or that the illness itself involves the possession of evil spirits. But, again, someone who believes such things should still be capable of entertaining the very same concept as we do – the PLAGUE. Indeed, it is necessary for them to have the very same concept in order to make sense of the idea that we can disagree with them

about the nature and cause of the disease. Ignorance and error are as problematic for the Theory Theory as they were for the Classical Theory.

Still, whether two people are employing the same concept or not[15] is a difficult question. We suppose that many theorists would claim that it's simply inappropriate to insist that the *very same concept* may occur despite a difference in surrounding beliefs. The alternative suggestion is that people need only have *similar concepts.* The idea is that differences in belief do yield distinct concepts, but this is not problematic because two concepts might still be similar enough in content that they would be subsumed by the same psychological generalizations – and perhaps that's all that really matters.

As tempting as this position may be, it is actually fraught with difficulty. The problem is that when the notion of content similarity is unpacked it generally presupposes a prior notion of content identity (Fodor and Lepore 1992). For example, a common strategy for measuring content similarity is in terms of the number of constituents that two concepts share. If they overlap in many of their constituents, then they are said to have similar contents (see, e.g., Smith et al. 1984). But notice that this proposal works only on the assumption that the shared, overlapping constituents are the same. So the notion of content similarity is illicitly building on the very notion it is supposed to replace.

Since the scope of this problem hasn't been absorbed in either philosophical or psychological circles, it pays to explore some other proposed solutions. Consider, for example, a suggestion by Eric Lormand (1996). Lormand claims that even a completely holistic theory of content needn't have any difficulties with stability; in other words, stability isn't supposed to be a problem even for a theory that claims that any change in the total belief system changes the content of every single belief. The trick to establishing stability, Lormand claims, is the idea that a given symbol has *multiple meanings.* Each of its meanings is given in terms of a subset of its causal/inferential links. Lormand calls these subsets *units* and asks us to think of a unit "as a separable rough test for the acceptable use of that representation" (1996: 57). The proposal, then, is that a holistic system of representation can allow for stability of content, since, as the system exhibits changes, some of a concept's meanings change, but some don't. To the extent that it keeps some of its units intact, it preserves those meanings.

Unfortunately, this suggestion doesn't work. Since Lormand's units are themselves representations, they are part of the holistic network that determines the content of every concept in the system. As a result, every concept embedded in any unit will change its meaning as the other meanings in the inferential network change. And if they change their meaning, they can't be the basis of the stability for other concepts (Margolis and Laurence 1998).

Paul Churchland (1998) has proposed a different solution. For some time, Churchland has been developing an approach to mental content known as *state-space semantics.* State-space semantics is a theory of content for neural networks where content is supposed to be holistic. To a first approximation, the content of an activation vector – i.e., a pattern of activation across an assembly of nodes in

such a network – is supposed to be determined by its position within the larger structure of the network. Since this position will be relative to the positions of many other nodes in the network, state-space semantics should have considerable difficulties in achieving content stability. As a result, Churchland is quick to reject content identity in favor of content similarity.

In earlier work, Churchland adopted a model much like the one in Smith et al. (1984). Imagine a connectionist network with a series of input nodes, output nodes, and an intermediary set of so-called hidden nodes. Taking the hidden nodes as specifying contentful dimensions, we can construct a semantic space of as many dimensions as there are hidden nodes, where points within the space correspond to patterns of activation across the hidden nodes. Supposing for simplicity that there are only three hidden nodes, the resulting semantic space would be a cube, each of whose axes corresponds to a particular hidden node and its level of activation. On Churchland's early treatments, content similarity was understood as relative closeness in a space of this sort. But this approach runs into much the same problem as the Smith et al. account. It only explains similarity of content by presupposing a prior notion of identity of content, one that applies to the constituting dimensions of the space.

In light of this difficulty, Churchland has recently put forward a new account of similarity of content. In the new model, Churchland suggests:

> A point in activation space acquires a specific semantic content not as a function of its position relative to the constituting axes of that space, but rather as a function of (1) its spatial position relative to all of the *other contentful points within that space*; and (2) its causal relations to stable and objective *macrofeatures of the external environment*. (1998: 8)

This new position, Churchland tells us, "constitute[s] a decisive answer to Fodor and Lepore's challenge" (ibid: 5) to provide a workable holistic account of content similarity.

Yet far from being a decisive answer to the challenge, Churchland's new account is really no improvement at all. His first determinant of content – spatial position relative to other contentful points in the space – immediately confronts a serious difficulty. Supposing that two networks do have nodes with the same overall relative positions, this alone doesn't suffice to fix their contents; one might well wonder why any given node in either network has the particular content it has (and not some other content). For example, Churchland describes one type of network as representing distinct families as it extracts four prototypical faces given photographs as input. But what makes it the case that the network's nodes represent families and faces as opposed to any of a wide variety of potential objects? In response to this problem, Churchland can only appeal to the resources of his second determinant of content – causal relations to features of the environment. The problem with this answer, however, is that this isn't a version of the Theory Theory at all. Rather, it relies on an atomistic theory of content of the

sort we discuss in the next section. The relation of the node to its surrounding nodes turns out to have nothing to do with its content; what matters for content is just the existence of a reliable causal link to features of the environment.[16] Of course, these reliable links provide stability, but that's because they underwrite a theory of content identity: Two nodes have identical contents just in case they are linked to the same environmental feature. So it's no surprise that Churchland can have a notion of similar content, since he helps himself to an independent account of sameness of content, despite his rhetoric to the contrary.[17]

Stability, it turns out, is a robust constraint on a theory of concepts. What this means for the Theory Theory is that mental theories make for bad cores. They have as much trouble as the Prototype Theory when it comes to reference, and they are especially bad in securing stability. If a version of the Dual Theory of concepts is to succeed, it looks like it's not going to be one whose cores have either classical structure or theory structure.

8.4 Concepts Without Structure

We've seen that the main views of conceptual structure are all problematic. In light of these difficulties, a number of theorists have proposed to explore the possibility that lexical concepts don't have any structure – a view known as *Conceptual Atomism* (see, e.g., Fodor 1998; Leslie 2000; Millikan 1998, 2000). Central to Conceptual Atomism is the thesis that a concept's content isn't determined by its relation to any other particular concepts. Instead, it's determined by a mind–world relation, that is, a causal or historical relation between the symbol and what it represents. Not surprisingly, Atomism finds its inspiration in Kripke's and Putnam's treatment of natural kind terms, only it's intended to cover a broader range of semantic items and is directed, in the first instance, to the nature of the conceptual system, not to language.

The most difficult task for an atomist is to provide a sufficiently detailed account of the mind–world relation that's supposed to determine conceptual content. One general strategy is to explain content in terms of the notion of co-variation (the same notion that we saw was illicitly at play in Churchland's treatment of stability). The idea is that a concept represents what it causally co-varies with. For example, if the concept D were tokened as a reliable causal consequence of the presence of dogs, then, on the present account, the symbol would express the property *dog* and be the concept DOG. Notice, however, that this simple account won't do. The reason is because all sorts of other things will reliably cause tokenings of the symbol D. This might happen, for example, as a result of perceptual error. On a dark night you might catch a fox out of the corner of your eye and mistake it for a dog running past your car.

Atomists have a number of resources for ruling out the non-dogs. One is to add the further condition that a concept represents what it would co-vary with

under *ideal conditions* (allowing for the possibility that non-dogs cause DOGS when the conditions aren't ideal; see, e.g., Stampe 1977; Fodor 1981/90). Another option is to say that a concept represents what it has the *function* of co-varying with (allowing for the possibility that the concept, or the system that produces it, isn't functioning properly in the non-dog cases; see, e.g., Dretske 1995; Millikan 1984, 1993). Yet another possibility is to say that the dog/DOG dependence is, in a sense, more basic than the non-dog-yet-dog-like/DOG dependence. For instance, the former dependence may hold whether or not the latter does, but not the other way around (Fodor 1990).

Though each of these strategies has its own difficulties, we want to focus on more general problems with Atomism, ones that aren't tied to the details of any particular atomistic theory. We'll mention three.

The first objection concerns the explanatory role of concepts. Most theories tie a concept's explanatory potential to its structure. This is evident in the other theories we've reviewed. For instance, the Prototype Theory explains a wide variety of psychological phenomena by reference to conceptual structure – categorization, typicality judgments, efficiency of use, and so on. The problem with Conceptual Atomism, however, is that it says that concepts have no structure. So it would seem that they can't really explain anything. Then what good are they?

The second objection is the worry that Conceptual Atomism is committed to an extremely implausible degree of innateness. In fact, Jerry Fodor, the most vocal defender of Atomism, has made this connection explicitly, defending the claim that virtually all lexical concepts are innate, including such unlikely candidates as CARBURETOR and QUARK. As Fodor sees it, the only way that a concept could be learned is via a process of construction, where it is assembled from its constituents. Since Atomism maintains that lexical concepts have no constituents, they must all be innate (Fodor 1981). But if CARBURETOR is innate, something has definitely gone wrong; maybe that something is Atomism itself.

The third objection is that atomistic theories individuate concepts too coarsely. Since they reduce content to a causal or historical relation between a representation and what it represents, concepts would seem to be no more finely individuated than the worldly items they pick out. Yet surely that isn't fine enough. The concept WATER isn't the same thing as the concept H_2O – someone could have the one without the other – but presumably they pick out the very same property. Or to take a more extreme case, the concept UNICORN isn't the same thing as the concept CENTAUR, yet because they are empty concepts, they would seem to pick out the very same thing, viz., nothing. So it's hard to see how an atomistic theory could tease such concepts apart.

Let's take these objections in reverse order. No doubt, the problem of achieving a fine-grained individuation is a serious concern for Atomism, but atomists do have a few resources they can call upon. For instance, in the case of empty concepts, they can maintain that the content determining co-variation relation is a nomic relation between properties. This helps because it's plausible there can be nomic relations between properties even if they are uninstantiated (Fodor 1990).

With other examples, atomists can distinguish co-referential concepts by insisting that one of the concepts is really complex and that its complexity isn't in dispute. Presumably, this is how they would handle the WATER/H_2O case – by maintaining that the concept H_2O incorporates, among other things, the concept HYDROGEN (Fodor 1990). Of course, there are other challenging cases for which neither of these strategies will work. Here we have in mind pairs of primitive concepts that express nomologically co-extensive properties (e.g., BUYING/SELLING, CHASING/FLEEING, EXTENDED/SHAPED). These prove to be the most difficult cases, since the natural solution for distinguishing them is to say they are associated with different content-determining inferences. Whether atomists have an alternative solution is very hard to say.

But let's turn to the other objections to Atomism, which, on the face of it, leave the atomist with even less room to maneuver. If Atomism says that lexical concepts have no structure, must they all be innate? And if lexical concepts have no structure, why aren't they explanatorily inert?

Fodor's argument for radical concept nativism has caused quite a stir in philosophy of mind, with theorists of different sorts dropping any doctrine thought to be tied up with the thesis.[18] As a result, the argument has not received the sort of careful critical scrutiny that it deserves. We believe that Atomism has been unfairly burdened with Fodor's strong nativist thesis, and that in fact it is possible to provide a satisfying account of how new primitive concepts can be acquired in a way that is compatible with Conceptual Atomism. The key here is the notion of a *sustaining mechanism*. Sustaining mechanisms are mechanisms that underwrite the mind–world relation that determines a concept's content. These will typically be inferential mechanisms of one sort or another, since people clearly lack transducers for most of the properties they can represent. Importantly, however, these inferential mechanisms needn't give rise to any analyticities or to a concept's having any semantic structure, since no particular inference is required for concept possession. Thus, such inferential mechanisms are fully compatible with Conceptual Atomism.

We are now in a position to see why Atomism is not committed to radical concept nativism. What the atomist ought to say is that the general question of how to acquire a concept should be framed in terms of the more refined question of how, given the correct theory of content, someone comes to be in a state of mind that satisfies the theory (Margolis 1998; Laurence and Margolis 2002). On an atomistic treatment of content this is to be understood in terms of the possession of a suitable sustaining mechanism. So the question of acquisition just is the question of how sustaining mechanisms are assembled. And here there are many things that an atomist can say, all consistent with the claim that concepts have no structure. For example, one type of sustaining mechanism that we've explored in detail supports the possession of natural kind concepts (see Margolis 1998; Laurence and Margolis, forthcoming). The model is based on what we call a *syndrome-based sustaining mechanism*, one that incorporates highly indicative perceptual information about a kind together with a disposition to treat something as a member of the

same kind so long as it shares the same constitutive hidden properties (and not necessarily the same perceptual properties) as the category's paradigmatic instances. The suggestion is that people have a general tendency to assemble syndrome-based sustaining mechanisms in accordance with their experience. Such a mechanism then establishes the mind–world relation that atomists say is constitutive of content, and together with environmental input is capable of delivering a wide range of unstructured concepts. Since the mechanism respects the character of one's experience – acquisition proceeds by the collection, storage, and manipulation of information to produce a representation that tracks things in the concept's extension – we think it is fair to say that this is a learning model.

Turning finally to the charge that Atomism leaves concepts explanatorily inert, the best strategy for the atomist is to say that the explanatory roles that are often accounted for by a concept's structure needn't actually be explained directly in terms of the concept's nature. The idea is that the atomist can appeal to information that happens to be associated with the concept; that is, the atomist can make use of the relations that a concept C bears to other concepts, even though these others aren't constitutive of C. This may seem a drastic step, but virtually any theory of concepts will do the same in order to explain at least some inferences in which concepts participate. Perhaps as a child you were frightened by a dog and as a result you've come to believe that dogs are dangerous. This belief may well explain quite a lot of your behavior toward dogs. Nonetheless, a classical theorist would not likely suppose that it was part of the definition of DOG that dogs are dangerous. All theories of concepts say that some of a concept's relations to other concepts are constitutive of its identity and some are not. And having made that distinction, it's sometimes going to be the case that how a concept is deployed will reflect its non-constitutive relations. The atomist simply takes this position to the limit and says that this is always the case. A concept's role in thought can't help but reflect its non-constitutive relations, since what's constitutive of a concept isn't its relation to any other particular concepts but just how it is causally (or historically) related to things in the world. One wonders, however, whether the atomist has gone too far. Could it really be that *none* of the ways in which a concept is deployed is explained by its nature?

8.5 Rethinking Conceptual structure

There's something unsettling about the claim that the explanatory functions of concepts are handled by their incidental relations. Consider once again typicality effects. Typicality effects are so pervasive and so rich in their psychological import that they constitute one of the central explananda of any theory of concepts. Indeed, it is largely because of the Classical Theory's failure to account for these effects that psychologists abandoned the Classical Theory in droves. Notice, however, that Conceptual Atomism is no different than the Classical Theory in its

capacity to deal with typicality effects. By maintaining that concepts have no structure, atomists are committed to the view that a concept's nature has no bearing whatsoever on its role in typicality effects. Of course, this doesn't mean that atomists have to deny the existence of typicality effects. Yet it is puzzling that some of the most important psychological data involving concepts end up having nothing at all to do with their nature.

At the same time, there are compelling pressures mitigating in favor of Atomism's central claim that concepts don't have any structure. In particular, all attempts to explain reference determination in terms of a concept's structure run into formidable difficulties. The Classical Theory, the Prototype Theory, and the Theory Theory all fall prey to the problems of ignorance and error, and each theory has its own peculiar difficulties as well.

The way out of this impasse lies in two related insights about conceptual structure that are implicit in the Dual Theory. The first of these is simply that concepts can have multiple structures. Thus in the original Dual Theory concepts were taken to have cores and identification procedures. The second insight is less obvious but it's really the crucial one. This is that concepts can have categorically different types of structure answering to very different explanatory functions.[19] The Dual Theory implicitly recognizes this possibility in the distinct motivations that it associates with cores and identification procedures. But once the point is made explicit, and once it is made in perfectly general terms, a whole new range of theoretical possibilities emerges.

The most immediate effect is the Dual Theory's recognition that the function of explaining reference may have to be teased apart from certain other functions of concepts. This would free the other types of structure that a concept has from a heavy burden and, crucially, would imply that not all conceptual structure is reference-determining structure. Having taken this step, one can then inquire about what other types of conceptual structure there are and about the specific functions they answer to.

We suggest that there are at least four central types of structure:

Compositional reference-determining structure This is structure that contributes to the content and reference of a concept via a compositional semantics. This type of structure is familiar from the Classical Theory. Whether any lexical concepts have this type of structure will depend on whether the problems of analyticity and ignorance and error can be met and whether definitions can actually be found. However, it is more or less uncontroversial that phrasal concepts such as BROWN DOG have this kind of structure. BROWN DOG is composed of BROWN and DOG and its reference is compositionally determined by the referential properties of its constituents: Something falls under BROWN DOG just in case it's brown and a dog.

Non-semantic structure This is structure that doesn't contribute to the content of a concept but does contribute significantly to some other theoretically

important explanatory function of concepts. Though the Dual Theory is not explicit about this, it seems plausible to think of Dual Theory's commitment to prototypes as a commitment to non-semantic structure.

Non-referential semantic structure This is structure that contributes to the content of a concept but is isolated from referential consequences. Though our discussion of the meaning or content of concepts has focused on their referential properties, these may well not exhaust the semantic properties that concepts possess. This type of structure would apply to, among other things, so-called narrow content.[20]

Sustaining mechanism structure This is structure that contributes to the content of a concept indirectly by figuring in a theoretically significant sustaining mechanism. Sustaining mechanism structure determines the referential properties of a concept, but not via a compositional semantics. Rather, this type of structure supports the mind–world relation that (directly) determines a concept's content.

These four different types of structure point to a range of new theoretical options that bear exploring. By way of illustration, we will briefly sketch a resolution to the impasse between Conceptual Atomism and the pressure to appeal to a concept's structure in explaining its most salient behavior.

If we look back at the Dual Theory, the main problems it faces center around its treatment of conceptual cores. We've seen that both definitional structure and theory structure are equally problematic in this regard. Neither is especially suited to reference determination; and, in any case, definitions have proven to be quite elusive, while theory structure has its difficulties with stability. Notice, however, that there is now an alternative account of cores available. Given the distinctions we have just drawn among the four types of conceptual structure, Conceptual Atomism is best construed not in terms of the global claim that lexical concepts have no structure at all, but rather as claiming that they have no *compositional reference-determining structure*. This opens the possibility that the cores of concepts might be atomic.

Indeed, atoms seem to be almost perfectly suited to fill the explanatory roles associated with conceptual cores. If cores are atomic, then one doesn't have to worry about the fact that concepts aren't definable. Atomism implies that they aren't. Similarly, if cores are atomic, then one doesn't have to worry about stability. Atomism implies that a concept's relations to other concepts can change as much as you like so long as the mind–world relation that determines reference remains in place. Atomic cores also explain the productivity of concepts: complex concepts are generated through the classical compositionality of atomic cores. The only explanatory role associated with cores that atoms seem to have trouble with is accounting for our most considered judgments about category membership. However, it's hardly clear that this is a legitimate desideratum for a theory of conceptual cores in the first place. If Quine's work on analyticity shows

anything, it's that people's most considered judgments of this sort are holistic, so it's not plausible to suppose that all of this information could be isolated for each concept taken individually. Dropping this last desideratum, then, there is a good case to be made for thinking that cores should be atomic.

At the same time, a model of this sort avoids the objection that Atomism is psychologically unexplanatory. We can agree with atomists that lexical concepts generally lack compositional reference-determining structure, but this doesn't mean we have to say that concepts are entirely unstructured. For example, prototypes and sustaining mechanisms may very well be part of a concept's structure. It's just that this structure doesn't directly determine its reference; reference is fixed by the mind–world relation that implicates cores, leaving prototypes (and other types of structure) to explain other things. And prototypes, for one, do explain many other things. Given their tremendous psychological significance, prototypes should be taken to be partly constitutive of concepts if anything is.

Concepts are psychological kinds. As we see it, the best theory of concepts is one that takes their psychological character seriously. The way to do this is to adopt a theory that admits different types of conceptual structure while tying them together by maintaining that concepts have atomic cores. In any event, it pays to focus on the nature of conceptual structure itself. Articulating the different explanatory roles for postulating conceptual structure and teasing these apart opens up a range of unexplored and potentially very promising theoretical options in the study of concepts.

Notes

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

1 This view of the nature of thought is not entirely uncontroversial. Yet it's difficult to see how finite creatures without access to a structured system of representation could be capable of entertaining the vast number of thoughts that humans have available to them. Even if we stick to relatively simple thoughts, the number of these is truly astronomical. For example, there are 10^{18} simple statements of sums involving numbers less than a million. This is more than the number of seconds since the beginning of the Universe and more than a million times the number of neurons in the human brain. How could a theory of thought accommodate these facts without postulating a structured representational system in which the same elements – concepts – can occur in different positions within a structured assembly? In any event, if a theory really says that thoughts don't have constituents, perhaps the best thing to say is that, according to that theory, there aren't any such things as concepts.

2 We will assume that thoughts and concepts have semantic properties and that chief among these are their truth-theoretic properties. We take it to be an important constraint on a theory of concepts that, e.g., the concept DOG refers to dogs.

3 Still, it is worth noting that the theories we discuss can be adapted with slight modification to alternative frameworks that take different stands on these foundational questions.

4 For more detailed surveys and development of the views here, see Laurence and Margolis (1999; in prep.). See also Smith and Medin (1981).

5 The main reason for the qualification is that, according to the Classical Theory, some concepts have to have no structure; these are the primitive concepts out of which all others are composed. Classical theorists have had little to say about how the reference of a primitive concept is fixed. But the most venerable account, owing to the British empiricists, is that primitive concepts express sensory properties and that they refer to these simply because they are causally linked to such properties via sensory transducers.

6 Work on the theory of concepts has become increasingly interdisciplinary, and many of the theories we will discuss bear the marks of ideas and motivations which have been transferred across disciplinary boundaries, particularly between psychology and philosophy. In line with much of this research, we take concepts to be mental representations (and thus mental particulars), since this perspective makes the most sense of the various psychological explananda that have rightly exerted considerable pressure on theorizing about concepts – even in philosophical circles. The reader should note that this is not a universally shared perspective and that many philosophers insist on construing concepts as abstract entities of one sort or another. Nonetheless, theorists who take concepts to be abstracta also take a deep interest in questions about conceptual structure. It's just that the structure in question is supposed to be the structure of abstract entities. See, e.g., Peacocke (1992) and Bealer (1982).

7 As the examples here indicate, the Classical Theory (and indeed all the theories we will be discussing) is, in the first instance, a theory about the nature of concepts that correspond to words in natural language – what are called *lexical concepts*. This is because theorists interested in concepts assume that the representations corresponding to natural language phrases or sentences are structured.

8 The motivation for the Classical Theory is by no means limited to these virtues. For example, another influential point in favor of this theory is its ability to explain our intuitions that certain statements or arguments are valid even though, on the face of it, they fail to express logical truths, e.g., "John is a bachelor, so John is unmarried" (see, e.g., Katz 1972).

9 Classical theorists have had little to say in defense of the notion of analyticity. E.g., Christopher Peacocke's seminal book on concepts (1992) falls squarely in the classical tradition, especially in its commitment to definitions, yet Peacocke takes little notice of the problems associated with analyticity, simply stating in a footnote that he is committed to some version of the analytic/synthetic distinction (see p. 244, fn 7). See Katz (1997), however, for a rare classical defense of analyticity, especially in the face of the present considerations.

10 In the most extreme cases, people know hardly any information at all. For instance, Putnam remarks that he can't distinguish elms from beeches, that for him they are both just trees. Yet arguably, he still has two distinct concepts that refer separately to elms and beeches. That wouldn't be possible if the mechanism of reference had to be an internalized definition.

11 What we are calling "the Prototype Theory" is an idealized version of a broad class of theories, one that abstracts from many differences of detail. This is true of each of the theories we present, though the diversity is perhaps more pronounced in the case of the Prototype Theory. For discussion of some of the different varieties, see Smith and Medin (1981).

12 The Dual Theory should not be confused with so-called *Two Factor* theories in philosophy. Though there are similarities, the Dual Theory and Two Factor theories address different issues. Two Factor theories are primarily concerned with distinguishing two different types, or aspects, of content. One factor accounts for all aspects of content that supervene on a person's body or that would be shared by molecule for molecule duplicates ("narrow content"). The other factor accounts for aspects of content that go beyond this, involving the person's relation to her environment ("wide content"). As a result, the two types of structure in the Dual Theory cross-classify the two aspects of content in Two Factor theories (see note 20 below).

13 According to the Theory Theory, the structure of a concept is constituted by its relations to the other concepts that are implicated in an embedding theory. Notice that on this account the structure of a concept can't be understood in terms of part/whole relations. For this reason, we have distinguished two models of conceptual structure (see Laurence and Margolis 1999). The first, the Containment Model, says that one concept, C_1, is included in the structure of another, C_2, just in case C_1 is literally contained in (i.e., is a proper part of) C_2. The second, the Inferential Model, says that C_1 is included in the structure of C_2 just in case C_1 stands in a privileged inferential relation to C_2. As should be evident from this characterization, the Theory Theory has to be construed in terms of the Inferential Model, but the Classical Theory and the Prototype Theory could be construed in terms of either model, depending on the exact motivations that support the postulation of classical and prototype structure.

14 These particular domains have been the subject of intense interdisciplinary investigation in recent years. For common-sense psychology, see Davies and Stone (1995a, 1995b), Carruthers (1996); for common-sense physics, see Spelke (1990), Baillargeon (1993), Xu and Carey (1996); for common-sense biology, see Medin and Atran (1999).

15 Or, for that matter, whether the same person is employing the same concept over time.

16 At best, Churchland's model shows how psychological *processes* could be holistic. They are holistic because they involve activation patterns across massively connected nodes in a network. But this doesn't mean that the *semantics* of the network are holistic.

17 It should be noted that Churchland is something of a moving target on these issues, though he often neglects to acknowledge changes in his view. For instance, in addition to the positions mentioned in the text, Churchland also tries maintaining that content similarity is a matter of similarity of "downstream processing" (see esp. 1996: 276),

> It is this downstream aspect of the vector's computational role that is so vitally important for reckoning sameness of cognitive content across individuals, or across cultures. A person or culture that discriminated kittens reliably enough from the environment, but treated them in absolutely every respect as a variant form of wharf-rat, must be ascribed some conception of "kitten" importantly different from our own. On the other hand, an alien person or species whose expectations of and behavior towards kittens precisely mirror our own must be ascribed the same concept "kitten," even though they might discriminate kittens principally by means of alien olfaction and high-frequency sonars beamed from their foreheads.

Apart from making his "state space semantics" have nothing whatsoever to do with the state space, this position falls prey to exactly the same sorts of problems as Churchland's first position, namely, it presupposes a notion of content identity for the "downstream" states that fix the content of the *kitten* vector.

18 See, e.g., Churchland (1986) and Putnam (1988).

19 These two points go hand in hand, since it's to be expected that if a concept has multiple structures that these would be of categorically different types.

20 The nature of narrow content is controversial but the main idea is that narrow content is shared by molecule-for-molecule duplicates even if they inhabit different environments. On some Two Factor theories (see note 12), a concept's narrow content is determined by its inferential role – a view that closely resembles the Theory Theory's account of conceptual structure. The difference is that, on a Two Factor theory, the inferential role of a concept isn't supposed to determine its reference.

References

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*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bealer, G. (1982). *Quality and Concept*. Oxford: Clarendon Press.

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*. New York: Cambridge University Press.

Carey, S. (1982). "Semantic Development: The State of the Art." In E. Wanner and L. Gleitman (eds.), *Language Acquisition: The State of the Art*. New York: Cambridge University Press.

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

Carruthers, P. (ed.) (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Churchland, P. M. (1996). "Fodor and Lepore: State-Space Semantics and Meaning Holism." In R. McCauley (ed.), *The Churchlands and Their Critics*. Cambridge, MA: Blackwell.

—— (1998). "Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered." *Journal of Philosophy*, XCV (1): 5–32.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: The MIT Press.

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

Davies, M. and Stone, T. (eds.) (1995a). *Folk Psychology*. Oxford: Blackwell.

—— (eds.) (1995b). *Mental Simulation*. Oxford: Blackwell.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Fodor, J. D., Fodor, J. A., and Garrett, M. (1975). "The Psychological Unreality of Semantic Representations." *Linguistic Inquiry*, 6: 515–32.

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

—— (1981/90). "Psychosemantics; or, Where Do Truth Conditions Come From?" In N. G. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell.

—— (1990). "A Theory of Content, II: The Theory." In *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
—— (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.
Fodor, J. A. and Lepore, E. (1992). *Holism: A Shopper's Guide*. Cambridge, MA: Basil Blackwell.
—— (1996). "The Red Herring and the Pet Fish: Why Concepts Still Can't Be Prototypes." *Cognition*, 58: 253–70.
Gettier, E. (1963). "Is Justified True Belief Knowledge?" *Analysis*, 23: 121–3.
Gopnik, A. and Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.
Katz, J. (1972). *Semantic Theory*. New York: Harper and Row.
—— (1997). "Analyticity, Necessity, and the Epistemology of Semantics." *Philosophy and Phenomenological Research*, LVII: 1–28.
Keil, F. (1994). "Explanation, Association, and the Acquisition of Word Meaning." In L. Gleitman and B. Landau (eds.), *The Acquisition of the Lexicon*, Cambridge, MA: MIT Press.
Kintsch, W. (1974). *The Representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Kripke, S. (1972/1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
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. and Margolis, E. (1999). "Concepts and Cognitive Science." In E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press.
—— (2002). "Radical Concept Nativism." *Cognition*, 86 (1): 22–55.
—— (in preparation). *The Building Blocks of Thought*.
Leslie, A. (2002). "How to Acquire a 'Representational Theory of Mind'." In D. Sperber and S. Davis (eds.), *Metarepresentations*. Vancouver Studies in Cognitive Science, vol. 10. Oxford: Oxford University Press.
Lormand, E. (1996). "How to Be a Meaning Holist." *Journal of Philosophy*, XCIII: 51–73.
Margolis, E. (1998). "How to Acquire a Concept." *Mind and Language*, 13 (3): 347–69.
Margolis, E., and Laurence, S. (1998). "Multiple Meanings and the Stability of Content." *Journal of Philosophy*, XCV (5): 255–63.
Medin, D. and Atran, S. (1999). *Folkbiology*. Cambridge, Mass: MIT Press.
Medin, D. and Ortony, A. (1989). "Psychological Essentialism." In S. Vosniadou and A. Ortony (eds.), *Similarity and Analogical Reasoning*. New York: Cambridge University Press.
Millikan, R. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
—— (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
—— (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.

——(2000). *On Clear and Confused Ideas: An Essay about Substance Concepts*. New York: Cambridge University Press.
Murphy, G. and Medin, D. (1985). "The Role of Theories in Conceptual Coherence." *Psychological Review*, 92 (3): 289–316.
Osherson, D. and Smith, E. (1981). "On the Adequacy of Prototype Theory as a Theory of Concepts." *Cognition*, 9: 35–58.
Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.
——(1998). "Implicit Conceptions, Understanding and Rationality." In E. Villanueva (ed.), *Philosophical Issues, 9: Concepts*. Atascadero, CA: Ridgeview Publishing Company.
Putnam, H. (1962). "The Analytic and the Synthetic." In H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science, Volume III*. Minneapolis: University of Minnesota Press.
——(1970). "Is Semantics Possible?" In H. Kiefer and M. Munitz (eds.), *Languages, Belief and Metaphysics*. New York: State University of New York Press: 50–63.
——(1975). "The Meaning of 'Meaning'." In K. Gunderson (ed.), *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.
——(1988). *Representation and Reality*. Cambridge, Mass.: MIT Press.
Quine, W. (1951/1980). "Two Dogmas of Empiricism." In *From a Logical Point of View: Nine Logico-Philosophical Essays*. Cambridge, MA: Harvard University Press: 20–46.
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*. Atlanta: Rodopi BV: 61–101.
Rosch, E. (1978). "Principles of Categorization." In E. Rosch and B. Lloyd (eds.), *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Rosch, E. and Mervis, C. (1975). "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology*, 7: 573–605.
Smith, E. (1995). "Concepts and Categorization." In E. Smith and D. Osherson (eds.), *Thinking: An Invitation to Cognitive Science, Vol. 3*. Second Edition. 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: 265–74.
Spelke, E. (1990). "Principles of Object Perception." *Cognitive Science*, 14: 29–56.
Stampe, D. (1977). "Towards a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy*. Minneapolis: University of Minnesota Press.
Wittgenstein, L. (1953/1958). *Philosophical Investigations*, trans. E. Anscombe. 3rd edition. Oxford: Blackwell.
Xu, F. and Carey, S. (1996). "Infants' Metaphysics: The Case of Numerical Identity." *Cognitive Psychology*, 30: 111–53.