ELSEVIER

Cognition 86 (2002) 25–55

COGNITION

www.elsevier.com/locate/cognit

Radical concept nativism ☆

Stephen Laurence[a], Eric Margolis[b,*]

[a]*Department of Philosophy and Hang Seng Centre for Cognitive Studies, University of Sheffield, Sheffield S10 2TN, UK*
[b]*Department of Philosophy and Rice Cognitive Sciences Program, MS-14, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA*

Received 9 May 2001; accepted 26 June 2002

Abstract

Radical concept nativism is the thesis that virtually all lexical concepts are innate. Notoriously endorsed by Jerry Fodor, radical concept nativism has had few supporters. However, it has proven difficult to say exactly what's wrong with Fodor's argument. We show that previous responses are inadequate on a number of grounds. Chief among these is that they typically do not achieve sufficient distance from Fodor's dialectic, and, as a result, they do not illuminate the central question of how new *primitive concepts* are acquired. To achieve a fully satisfactory response to Fodor's argument, one has to juxtapose questions about conceptual content with questions about cognitive development. To this end, we formulate a general schema for thinking about how concepts are acquired and then present a detailed illustration.

Keywords: Concept; Conceptual development; Concept nativism

1. Introduction

Noam Chomsky has argued that, contrary to empiricist doctrine, the real difficulty in accounting for cognitive capacities such as language is one of postulating a sufficiently rich innate mental endowment. Were we to limit ourselves to the methodological constraints of empiricism, we simply wouldn't be able to explain how children rapidly develop these capacities in such a uniform manner across widely varying and impoverished environments. While hardly uncontroversial, Chomsky's forceful case for nativist approaches to language has had a liberating effect. Theorists working on disparate areas of the mind now feel free to explore the possibility of developing strongly nativist models

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

* Corresponding author. Department of Philosophy, MS-14, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA. Fax: +1-713-384-5847.

E-mail address: margolis@ruf.rice.edu (E. Margolis).

0010-0277/02/$ - see front matter
PII: S0010-0277(02)00127-0

where a generation or two ago the prevailing climate would have made such models unthinkable. Still, even within this nativist-friendly climate, it is possible to go too far. Just about everyone thinks that this is exactly what Jerry Fodor has done. He has argued for the extraordinary claim that virtually all lexical concepts (concepts like CAT, CARBURETOR, and BROCCOLI) must be innate. According to Fodor, only patently complex phrasal concepts (concepts like BIG BLACK CAT) can be learned (Fodor, 1975, 1981).

Not surprisingly, Fodor has had few supporters.[1] Philosophers seem to have taken the conclusion to be so patently absurd that they think the argument behind it barely needs to be addressed. Hilary Putnam, for instance, dismisses the thesis as incompatible with the theory of evolution, while giving no diagnosis of where Fodor's argument actually goes wrong. His entire discussion – scarcely longer than the subheading that precedes it – is as follows (Putnam, 1988, p. 15):

> To have given us an innate stock of notions which includes *carburetor*, *bureaucrat*, *quantum potential*, etc., as required by Fodor's version of the Innateness Hypothesis, evolution would have had to be able to anticipate all the contingencies of future physical and cultural environments. Obviously it didn't and couldn't do this.[2]

Patricia Churchland's treatment is equally brusque. Speaking of Fodor's conclusion, she remarks that it is "difficult to take such an idea seriously" (Churchland, 1986, p. 389). And without offering any analysis of his argument, she states that she considers it to be a *reductio ad absurdum* of Fodor's Language of Thought Hypothesis.

As will become clear, we think that these reactions are deeply problematic. Apart from anything else, responses like these have encouraged a superficial understanding of Fodor's argument. This is unfortunate since, in spite of the near universal rejection of its conclusion, the dialectic that Fodor's argument generates remains extremely influential. In cognitive science, a number of theorists explicitly endorse the logic of his argument – though not its conclusion – and on this basis motivate rich and substantial research programs. For example, in the introduction to an important volume on lexical semantics, Beth Levin and Steven Pinker locate much of the impetus for research into lexical semantic structure by reference to Fodor's argument. They write (Levin & Pinker, 1991, 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

[1] Interestingly, however, Chomsky himself has expressed considerable sympathy for Fodor's position (Chomsky, 1991, p. 29):

"[C]hildren acquire knowledge of lexical items on the basis of very few presentations, perhaps only one, and under quite ambiguous circumstances. Furthermore, this is shared knowledge; children proceed in essentially the same way, placing the lexical entries in the same fixed nexus of thematic and other relations and assigning them their apparently specific properties. Barring miracles, this means that the concepts must be essentially available prior to experience, in something like their full intricacy. Children must be basically acquiring labels for concepts they already have, a view advanced most strongly by Jerry Fodor, and are somehow endowed with the capacity to identify the use of these concepts in real life situations…"

[2] Here's the subheading: "Our Concepts Depend on Our Physical and Social Environment in a Way That Evolution (Which Was Complete, for Our Brains, about 30,000 Year Ago) Couldn't Foresee" (p. 15).

> 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 a similar context, the linguist and cognitive scientist Ray Jackendoff says that he endorses the logic of Fodor's argument "unconditionally" (Jackendoff, 1989, p. 50).[3] But even among those who do not recognize it, Fodor's argument exerts a powerful influence on the field, often compelling the rejection of one or more theses about the mind for fear of being committed to a nativism as strong as Fodor's. In this way, his argument has had as much an effect on thinkers like Churchland as on Levin, Pinker, and Jackendoff.

We believe that Fodor's argument presents an important challenge to theories of concepts and should be viewed in much the same spirit as the Nelson Goodman's (1954) new riddle of induction, the W.V.O. Quine's (1960) indeterminacy thesis, and the skeptical paradox that Saul Kripke (1982) associates with Wittgenstein. Just as any adequate treatment of induction, meaning, or rule following must ultimately come to terms with these foundational challenges, so any adequate theory of the nature of concepts must ultimately come to terms with what might be called *Fodor's Puzzle of Concept Acquisition*. Similarly, just as these other philosophical puzzles are deeply bound up with the nature of rule following, meaning, and justification, so Fodor's Puzzle is inextricably bound up with the nature of concepts.[4]

In this paper we offer a comprehensive analysis and evaluation of Fodor's Puzzle, locating a general solution that ties issues about the nature of conceptual content directly to the question of how concepts are acquired. In the end, Fodor's Puzzle amounts to the challenge of explaining how a primitive or unstructured concept can be learned. We show that this challenge can be met but only by examining particular theories of content from a developmental perspective. We illustrate our solution with a detailed look at one sample theory of content. As we see it, this exercise has far-reaching implications. It shows that theories of content and theories of concept acquisition have to be juxtaposed in a way that

[3] Jackendoff is able to do this because, while he agrees with the logic of Fodor's argument, he disagrees in his assessment of the relevant empirical facts. In particular, he thinks that Fodor is wrong about the question of whether lexical concepts have internal structure. The significance of this disagreement will become clear in the next section.

[4] Of course, Fodor may not conceive of his argument as presenting a "puzzle" to be overcome, but his intentions aren't relevant to how other theorists should view the matter. This is no different than Quine's indeterminacy thesis or Kripke's skeptical paradox, which may be viewed (depending on one's theoretical predilections) as implausible yet established theses about meaning and rule following or, instead, as counterintuitive puzzles that ought to have adequate solutions.

hasn't been fully recognized. It is one of the chief virtues of Fodor's Puzzle that it highlights the significance of this neglect.

2. Fodor's argument that virtually all lexical concepts are innate

As Fodor uses the term, *concepts* are sub-sentential mental representations, that is, representations with sub-propositional contents.[5] Concepts in this sense are the building blocks of thought. It is because we have the concept of CHOCOLATE, for example, that we can think about chocolate, that we can desire chocolate, and that we can purposefully go about getting ourselves some chocolate. As with expressions in natural languages, some concepts are naturally understood to be composed of simpler elements. Just as the expression "huge pile of chocolate" is composed of the words "huge", "pile", etc., so the concept HUGE PILE OF CHOCOLATE can be understood to be composed of the concepts HUGE, PILE, etc. The concepts that correspond to natural language words (or morphemes) – i.e., lexical concepts – are the target of Fodor's discussion.[6] The conclusion that Fodor argues for is that virtually *all* lexical concepts are innate. Thus, according to Fodor, not only are concepts like CAUSE, AGENT and EDGE innate, but so too are MODEM, PLANET and CARBURETOR. Indeed, since normal adults command a vocabulary of at least 60,000 words,[7] it would seem that, at a bare minimum, they possess 60,000 innate concepts. Of course, on Fodor's view, the actual number of innate concepts would have to be far larger, since it has to do justice to the full range of potentially available lexical items. A better estimate might come from the number of words in the OED (half a million or so, according to Fodor). Even this, however, is likely to be a rather conservative estimate, since not all concepts have natural language correlates and new words are added to languages all the time, especially terms for cultural and technical innovations ("modem", "quark", "yuppy", and so on). Fodor's thesis, then, is bracingly strong. Accordingly, we will refer to the position as *radical concept nativism*.

In spite of the highly counterintuitive conclusion he reaches, Fodor's argumentative strategy is actually rather sensible. He begins by noting that not all forms of concept acquisition count as learning. For example, acquiring a concept as a result of a blow to the head isn't concept learning, nor is acquiring a concept as a result of taking high doses

[5] What concepts are is a matter of considerable dispute. However, it is doubtful that any of the controversy surrounding the nature of concepts affects the central issues concerning Fodor's argument.

[6] It is possible to distinguish several theoretically interesting categories that more or less coincide with the pre-theoretic notion of a word. One ties words to morphemes, which are traditionally characterized as the smallest units of language that are assigned semantic values (e.g., "unforgetable" is composed of three morphemes – "un", "forget", and "able"). Another conception treats words as *syntactic* atoms (making "unforgetable" a single word). See Di Sciullo and Williams (1987) for further discussion. The notion of a word that's at issue in Fodor's Puzzle is closer to the morphemic understanding than the syntactic one. While one could try to be more precise, this is plenty accurate enough for a first pass, which is all that seems to be needed to make sense of Fodor's somewhat vague claim that "most lexical concepts are innate".

[7] Pinker (1994, p. 150). The estimate is based on the work of William Nagy and Richard Anderson using American high school graduates. It is worth noting that this fact entails a truly remarkable rate of lexical acquisition (and hence concept acquisition). Averaging the acquisition over the course of 17 years, it works out to ten words (and the corresponding concepts) a day, or as Pinker notes, *one every hour and a half*.

of vitamins, or undergoing some futuristic neurosurgery. So learning models of concept acquisition need to be distinguished from other models. What seems to mark the cases of acquisition without learning is that the mechanism responsible for the acquisition in these cases is singularly non-rational. But, Fodor argues, rational acquisitional models must involve some form of hypothesis testing and confirmation.[8] And any non-trivial cases of concept acquisition by hypothesis testing must involve the construction of the acquired concept from primitive concepts. As a result, empiricism requires that lexical concepts are, in general, constructions, composed from simpler and ultimately primitive concepts. For Fodor, this is the point at which empiricism breaks down. The problem, he claims, is that the vast majority of lexical concepts are not definable in terms of more primitive concepts and there are no viable decompositional alternatives to definitions.

Fodor illustrates his argument with an example of a typical "concept learning experiment". In these experiments, the experimenter has a particular concept in mind, which is labeled with a novel predicate, say, "flurg". Subjects are asked to sort various stimuli according to whether they are flurg or not, where all they have to go on is the feedback that the experimenter provides after each trial. For example, if FLURG is the concept GREEN, then when the subject says that a card with a green circle on it is flurg, she'll be told that she is right. And if she says that a card with a red circle on it is flurg, she'll be told that she is wrong. And so on. Eventually, if all goes right, she'll come to reliably sort cards according to whether they are flurg.

While many of the experiments along these lines have been conducted by people with behaviorist leanings, it is hard to see that anything but a mentalistic interpretation of the concept learning task makes sense. In particular, subjects in the task appear to be employing an inductive procedure; that is, they frame hypotheses to themselves about the salient concept and compare these to the available data. Thus, our hypothetical subject might initially form the hypothesis that FLURG is the concept CIRCLE, and so she would take it as evidence in favor of her hypothesis that the experimenter affirms, in one or more trials, that cards with circles on them are flurg. Success comes when she corrects this error and finally settles on the hypothesis that the concept FLURG is that concept which satisfies the individuating conditions of GREEN.[9]

Now, as Fodor points out, there is clearly something odd about accounts of this kind. In order to frame and test hypotheses, one has to already possess the concepts in which they are couched, as well as the concepts that are necessary for tracking the evidence that bears upon them. For example, our subject has to be able to employ GREEN in order to represent both the critical hypothesis and the data that supports it. In what sense, then, does she come to *learn* the concept GREEN? At times, Fodor is tempted to conclude that these considerations show that concept learning simply isn't possible. But there is another way of looking at the situation, one that Fodor himself suggests as a way of salvaging hypothesis-testing models. This is to maintain that learned concepts have internal structure, that is, that they are composed of other (more primitive) concepts. Notice that so long as the target concept

[8] Indeed, he claims that these are the only *conceivable* models of rational concept acquisition. See Fodor (1975, p. 36).

[9] As this neutral formulation makes clear, nothing turns on the differences that arise between an extensional and an intensional semantics. See Fodor (1975, p. 95).

is complex, there is no need for its prior possession in order to represent its evidential base. With a complex concept, one can appeal to evidence that is framed just in terms of its constituents.[10] Similarly, there is no need for the prior possession of the concept in order to represent the critical hypothesis. With a complex concept, it can be assembled in the course of the hypothesis-testing procedure. The result, as Fodor sees it, is that if there is any sense to be made of concept learning, only complex concepts can be learned. Concepts with no internal structure aren't even candidates.

Given all this, empiricism stands or falls with the question of whether most lexical concepts are structured.[11] As Fodor puts it (Fodor, 1981, pp. 278–279):

> Roughly, what Empiricists and Nativists disagree about is the structure of lexical concepts. For the empiricist, lexical concepts normally have internal structure. … In particular, on the assumption that only sensory concepts are primitive… it must be that concepts like TRIANGLE, BACHELOR, XYLOPHONE, CHICAGO, HAND, HOUSE, HORSE, ELECTRON, GRANDMOTHER, CIGAR, TOMORROW, etc. are all internally complex. The empiricist must hold this because, by stipulation, the empiricist view is that the attainment of non-sensory concepts involves learning the truth of a hypothesis that exhibits their internal structure.

Fodor goes on to argue that the only relevant kind of internal structure that a concept can have is definitional structure. He doesn't offer a principled reason for thinking this. Rather, he just stresses that there aren't any serious alternatives. In particular, he argues that the Prototype Theory is a nonstarter, despite the fact that it continues to command widespread support in cognitive science.[12] According to Fodor, concepts couldn't be prototypes because compositionality is essential to conceptual systems, and prototypes don't compose. First, many complex concepts don't have a prototype. To use one of Fodor's examples, though there may be a prototype for GRANDMOTHER, there's no prototype for GRANDMOTHERS MOST OF WHOSE GRANDCHILDREN ARE MARRIED TO DENTISTS (Fodor, 1981, p. 297). Second, when complex concepts do have prototypes, they still needn't be inherited from their component concepts. For example, the prototype for PET FISH (which picks out things like goldfish) makes little contact with the prototypes associated with PET (which picks out dogs, cats, and so on) or with FISH (which picks out something more like a trout). For Fodor, given the centrality of compositionality to the conceptual system, these gross failures of compositionality suggest that prototypes simply do not capture what is essential to concepts.

Having ruled out non-definitional accounts of internal structure, the issue now turns on whether lexical concepts are definable. Fodor's primary argument here is that there just

[10] For example, the concept GREEN OR SQUARE can be learned on the basis of the evidence that x is flurg and x is green, y is flurg and y is square, etc. (without any representation of green squares as such).

[11] It's not just empiricism that's at stake. If the argument so far is correct, more moderate forms of nativism than Fodor's turn on exactly the same question.

[12] According to the Prototype Theory, concepts have statistical structure. One way of understanding this claim is that a complex concept *C* has prototype structure if its constituents express properties that things that fall under *C* tend to have. For elaboration and critical discussion, see Laurence and Margolis (1999) and Smith and Medin (1981).

don't seem to be any definitions.[13] He's impressed by what he sees as a long history of failed philosophical projects attempting to analyze such philosophical concepts as JUSTICE and KNOWLEDGE. And following Wittgenstein, he argues that ordinary concepts fare no better. Where Wittgenstein famously argued against the possibility of defining a concept as apparently simple as GAME, Fodor (1981) considers several proposals for the concept PAINT_{tr}, corresponding to the transitive verb "paint". His example is especially dramatic since he claims that PAINT_{tr} cannot be defined even using, among other things, the concept PAINT (corresponding to the noun "paint").

The first definition Fodor considers is: X COVERS Y WITH PAINT (based on Miller, 1978). He 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 with paint, this doesn't count as an instance of PAINTING. The factory or the explosion doesn't paint the spectators, yet the case satisfies the proposed definition. What seems to be missing is that an agent needs to be involved, and the surface that gets covered with paint does so as a result of the actions of the agent. In other words: X $\text{PAINTS}_{\text{tr}}$ 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. As Fodor points out, if you, an agent, kick over a bucket of paint and thereby cover your new shoes with paint, you haven't painted them. What seems to be needed is 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 $\text{PAINTS}_{\text{tr}}$ 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.[14]

The argument we end up with, then, is this.

Fodor's argument for radical concept nativism

1. Apart from miracles or futuristic super-science all concepts are either learned or innate.
2. If they're learned, they are acquired by hypothesis testing.

[13] He also argues that the few initially plausible candidate analyses that have been offered have never been corroborated in any psychological studies. See Fodor, Fodor, and Garrett (1975) and Fodor, Garrett, Walker, and Parkes (1980).

[14] To be fair, Fodor's discussion may not do justice to the advocate of definitions. In particular, it is 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.

(*continued*)

3.	If they're acquired by (non-trivial) hypothesis testing, they're structured.
4.	Lexical concepts aren't structured.
5.	So lexical concepts aren't acquired by hypothesis testing.
6.	So lexical concepts aren't learned.
7.	Therefore, lexical concepts are innate.

3. Some responses to Fodor's argument

Philosophers and cognitive scientists have raised a wide variety of objections to Fodor's argument. For the most part, these cluster around two general sorts of reactions. The first, which is especially prominent among philosophers, is to dismiss Fodor's argument on the basis of a direct assault on Fodor's conclusion or a quick counterexample to one of his premises. The second general reaction, which is more prominent among linguists and psychologists, accepts the overall logic of the argument but challenges Fodor's assessment of the empirical evidence against prototypes and definitions. Before we turn to our own response to Fodor's argument, we would like to say a bit about these other approaches.

We've already mentioned two attempts by philosophers to preempt Fodor's argument – Putnam's charge that Fodor's thesis is incompatible with evolution, and Churchland's suggestion that Fodor's thesis constitutes a *reductio ad absurdum* of the Language of Thought Hypothesis (see Section 1). Putnam and Churchland are not alone in endorsing these arguments. Kim Sterelny (1989), among others, gives a version of the argument from evolution, and Andy Clark (1994), while not going quite so far as Churchland, also takes Fodor's argument to constitute part of a strong *prima facie* case against the Language of Thought Hypothesis. Other philosophers have offered counterexamples to particular premises of Fodor's argument. Some have suggested that not all learning is a matter of hypothesis testing. For example, Jerry Samet and Owen Flanagan (Samet & Flanagan, 1989) note that food aversions occur on the basis of a single trial and also cite learning to swing a golf club; both of these are supposed to be cases of learning which don't involve hypothesis testing (see Sterelny, 1989 for similar examples). Samet (1986) also notes that there are cases that seem to involve acquisition in some sense, where what is acquired is neither learned nor innate. His cases aren't cognitive, but do involve rather suggestive analogies. One of these is that people acquire diseases – they catch colds, for instance – but that, in doing so, they neither learn to have the disease, nor do they possess it innately. Similarly, a camera might be said to "catch" pictures, which again, are neither learned nor innately built into the machine.

Unfortunately, none of these objections gets very far, and some, it turns out, are positively misleading. In spite of the eminent support that the objection from evolution has attracted, it is actually a very weak response to Fodor's argument. The main problem is that the objection is simply unilluminating, since it does nothing to pinpoint where Fodor's argument goes wrong. At best, the argument comes down to something like this: we have much better reasons to accept current accounts of evolution than we do for accepting Fodor's radical concept nativism, so given a conflict we should abandon the nativism. While this strategy isn't wholly without merit, it doesn't really do anything to resolve the

puzzle that Fodor raises. Not only does it leave unanswered all of the important questions that Fodor's argument raises for theories of concepts, it suggests that we should just *ignore* them. Of course, one could adopt this position, and leave it at that. After all, if it really does come down to a choice between Darwin and Fodor, Darwin wins hands down. The problem, however, is that this response really is just intellectually philistine. The point of philosophical puzzles isn't necessarily to accept their conclusions. Zeno doesn't convince us that we can't ever get across the room, and Goodman doesn't convince us that we really don't have any justification for thinking that newly discovered emeralds will be green. The point of these puzzles is that they seem to embody deep difficulties that infect our total theory of the world, puzzles about how we understand space and time, justification, ontology, meaning, etc. The value of such puzzles is exactly that they capture these difficulties, while providing a focused point of reflection. To simply side-step the problems they raise is to opt out of doing philosophy.

Similar points apply to Churchland's claim that Fodor's argument constitutes a *reductio ad absurdum* of the Language of Thought Hypothesis. Like the argument from evolution, Churchland's response fails to pinpoint what's wrong with Fodor's argument and sheds no light on the deeper issues about the nature of concepts that are connected with the argument. In Churchland's case, however, the summary treatment of Fodor's Puzzle isn't merely anti-philosophical. It is also deeply problematic. This is because, as it turns out, there is a fully satisfying way of answering Fodor's argument that does not require an abandonment of the Language of Thought Hypothesis.[15] So rejecting the hypothesis on these grounds is simply mistaken. Perhaps Churchland is right that thinking doesn't occur in a language-like system, but her response to Fodor's Puzzle offers no support for that position.

Now the examples of learning that don't involve hypothesis testing do attempt to locate where Fodor's argument goes wrong, essentially challenging premise 2. But they are also unilluminating, since these other forms of learning don't seem to apply to concepts – or at least, if they do apply to concepts, it is not at all clear how. Much the same goes for Samet's analogies, which seem to be aimed at premise 1. Again, it is not the least bit clear how the story is supposed to go for concept acquisition. Moreover, none of these responses, or those of Putnam and Churchland, provide any clues about the process of concept acquisition. They all effectively leave us with *no idea whatsoever how concepts are acquired*.

A more promising response along these general lines might challenge the representativeness of the so-called concept learning experiment that Fodor cites to illustrate his argument. One may wonder whether the concept learning experiment really does do justice to empiricist thought on how concepts are acquired. This is, we think, a substantive difficulty with the version of Fodor's argument we have presented. One problem is that not

[15] See the next section. For now, we might note that there have been many empiricist versions of the Language of Thought Hypothesis, so it is not at all clear that it is the Language of Thought Hypothesis that is at fault. It is also worth pointing out that our presentation of Fodor's argument makes no mention of the Language of Thought Hypothesis. Though the situation is complex, non-language-of-thought approaches, including connectionist approaches, don't have any special advantages for responding to Fodor's Puzzle. In fact, connectionist models don't even fall into a single category in the present context. Some are subject to Fodor's arguments against prototypes and definitions (e.g. theories that are invested in semantic reduction via "microfeatures", which aren't relevantly different from non-connectionist reductive accounts). Others are compatible with one or another of the solutions we discuss below.

all brands of concept acquisition that have the empiricist stamp of approval even count as learning. For the British Empiricists, the faculty of imagination is another major source of acquisition. David Hume, for example, says in the *Treatise on Human Nature*, "wherever the imagination perceives a difference among ideas, it can easily produce a separation", and these simple ideas "may be united again [by the imagination] in what form it pleases" (Hume, 1739/1978, p. 10). This is how we can get ideas of "winged horses, fiery dragons and monstrous giants" (Hume, 1739/1978, p. 10). These concepts are not learned at all, but neither are they innate, and that's a good part of what really matters to empiricists.[16] To make matters worse, it isn't even clear that typical empiricist models of concept *learning* should be thought of as involving hypothesis testing. Suppose in an empiricist vein that most concepts are complex representations that are assembled from their constituents in a way that reflects the environmental correlations that are detected by a sensitive statistical mechanism. Models of this sort are readily imaginable where the resulting concept is constructed without any hypotheses being put forward for confirmation. So the concept learning experiment that Fodor uses to motivate his argument doesn't seem to be particularly representative of classical empiricist accounts of concept acquisition.

Unfortunately, none of this really affects the dialectic very much. Fodor's argument can be easily reformulated to avoid these objections, and doing so helps to bring out what's really crucial to Fodor's case for radical concept nativism:

A revised version of Fodor's argument for radical concept nativism

1.	Apart from miracles or futuristic super-science all concepts are either constructed from primitives or innate.
2.	If they're constructed from primitives, they're structured.
3.	Lexical concepts aren't structured.
4.	So lexical concepts aren't constructed from primitives.
5.	Therefore, lexical concepts are innate.

Fodor's argument really turns on the issue of *conceptual structure*, and that isn't affected by skepticism about the significance of hypothesis testing.

Finally, we turn to the second sort of general response to Fodor's argument – the challenges to Fodor's critique of definitions and prototypes. These challenges are often accompanied by substantial empirical research programs. However, we will not be examining in detail any response to Fodor's argument that runs along these lines. This is not because we think that the empirical studies are uninteresting – far from it. But even granting the significance of these studies for a full understanding of concepts, it is important to appreciate that they don't address the fundamental issue raised by Fodor's Puzzle. As we see it – and as will become clearer in the next section – the core question that Fodor's Puzzle raises is whether one can acquire *new primitive concepts*. This is the heart

[16] A similar point applies in the case of phrasal concepts. Generally, these are neither learned on the basis of hypothesis testing nor innately given as such (see Samet & Flanagan, 1989).

of the matter because it focuses on the possibility of expanding the combinatorial expressive power of one's internal system of representation. Linguists and psychologists who challenge Fodor's empirical critique of definitions and prototypes more or less concede that such an expansion can't take place through a learning process (see again the quotes by Levin and Pinker, and Jackendoff in Section 1). Learning, for these linguists and psychologists, can only be what Fodor says it is, namely, a constructive process that operates on previously available innate primitive concepts. What we show in the next section is that this is a deeply misguided picture of the mind.

4. The proper treatment of Fodor's Puzzle

The key to understanding Fodor's Puzzle is seeing that it gains much of its plausibility from an extremely natural yet ultimately erroneous conception of the nature of concepts and how they are acquired. The puzzle is a vestige of reductive models of concepts and the delimiting range of options that they appear to offer when it comes to conceptual development. On a reductive model, concepts are taken to be either primitive or complex, and development consists of the construction of new complex concepts from previously available concepts. Under the classical version of this picture, complexes were taken to embody definitions, so development was understood in terms of the construction of a concept by articulating its definition. Other reductive theories, especially the prototype theory, may have relaxed the constraints that a complex concept bears to its constituents, but the vision of development has remained much the same. Acquiring a concept means assembling a complex concept. So powerful is this picture that alternatives may seem inconceivable. It is this apparent lack of options that Fodor's argument relies on. Fodor's argument turns on precisely the thought that concept learning requires construction from the primitives into which concepts decompose semantically, that there couldn't be a learning model for acquiring new primitive concepts. A good part of our response to Fodor is that this claim is simply wrong. Contrary to the shared assumption of both Fodor and his dialectical opponents, primitive concepts can be learned.

We begin by setting out the logical structure of our response to Fodor's Puzzle. As we see it, Fodor goes wrong because he fails to pose the issue of concept acquisition in its most fundamental terms. If possessing a concept means possessing a contentful representation, the issue of acquisition should be recast as the following question: *Given the correct theory of mental content, how can one come to be in a state in which the conditions that the theory specifies obtain?* For an adequate answer to this question, you need to look at the acquisition process from the vantage point of developed theories of content.[17] For philo-

[17] Several authors have independently hit upon this element of our general strategy (see, e.g., Block, 1986; Cowie, 1999; Margolis, 1995, 1998; Sterelny, 1989). The suggestion, however, has usually been understood in a skewed way, where it is supposed to be a virtue of the author's preferred theory of content that it can address Fodor's Puzzle, not that there is a general strategy that can be applied to a wide range of theories of content. In contrast, following Margolis (1995), we wish to characterize the strategy in its most general terms, to make clear that there is *schema* for answering Fodor's Puzzle. Of course, this isn't to say that all theories of content will be equally satisfying in their prospects for solving the puzzle. In fact, we believe that the ability to handle concept acquisition is an important measure of the fruitfulness of different theories of content. See the discussion of Block (1986) in Section 5.

sophers, this means getting your hands a bit dirty. You can't talk in the abstract about the question of whether specific concepts are learned or innate. Instead, you have to actually pick a theory of content and see how an acquisitional process might look in light of the constraints it imposes. Any theory of content that applies to primitive concepts will potentially offer a model for expanding the combinatorial expressive power of the conceptual system. Whether it does so depends on the details of the theory and the facts about how people might get into the sorts of states that the theory specifies.

To see how this general strategy plays out in concrete terms, we will discuss how primitive concepts might be acquired under a specific theory of content. We want to emphasize, however, that the specific theory of content and the particular account of acquisition that we discuss are simply illustrations of our strategy for addressing Fodor.

The sample theory of content which we will use is Fodor's own theory (Fodor, 1990a,b). Ironically, the theory of content that he has developed in the last 15 years or so has all the resources to provide a fully satisfactory answer to his puzzle of concept acquisition. Fodor's theory is useful since it clearly applies to primitive concepts – indeed, it was constructed specifically to accommodate Fodor's view that the vast majority of lexical concepts have no internal structure. For Fodor, having a concept is not a matter of knowing its definition or having a prototype. It is a matter of having a representation that stands in an appropriate mind–world relation. Fodor thinks that this relation is a specific type of causal relation. His account, the *asymmetric dependence theory*, is this: a mental representation expresses a property, say, the property doghood, in virtue of the fact that there is a nomic connection between doghood and DOG tokenings and the fact that, whenever there is a nomic connection between some other property and DOG tokenings, this other nomic connection is asymmetrically dependent on the dog/DOG connection. The latter condition is meant to rule out cases of error and other cases where a concept is caused by something that isn't in its extension. Fodor's idea is that circumstances like these wouldn't cause you to token DOG unless dogs caused you to token DOG, but that the reverse isn't true. The priority of the dog/DOG dependence is then supposed to explain why only dogs fall under the concept DOG.[18]

One unusual feature of this theory is that no specific piece of information that people associate with dogs via the concept DOG is actually constitutive of the concept. Fodor doesn't deny that our concepts are often the locus of significant bodies of information, but in an important sense none of this information is essential. In particular, all that matters to a concept's content are the dependency relations that it bears to things in the world. Of course, no one, not even Fodor, thinks that the information that is associated with a concept is completely irrelevant. There must be a reason why the mind–world relation obtains. And since how and when a concept is deployed is usually a function of the

[18] A standard example of an error would be a perceptual misidentification. For example, you might see a large cat dart in front of your car on a dark night and think that it is a dog. Since under these conditions you are disposed to token DOG in the presence of cats, a dependence exists between cats and DOG. The asymmetric dependence condition explains why, despite this fact, cats are *not* in the extension of DOG. The need to accommodate error is an important constraint on theories of content. Notice, for example, that if the theory of content in question were simply that categorization determines content – i.e., the view that something falls under the concept *C* just in case people apply *C* to it – then one could never *mis*apply a concept. But clearly error is a fact of human cognitive life. A theory of content that implies the contrary can be rejected out of hand.

information associated with it, it looks like Fodor is going to have to say that the asymmetric dependence relation itself depends upon the information that is associated with a concept. What Fodor doesn't have to accept, however, is that any particular mechanism of this sort – any particular belief or inference – is required for the possession of a given concept. As a result, it is perfectly compatible with Fodor's theory – in fact, it is one of its chief strengths – that different people can associate vastly different information with a concept yet nevertheless possess exactly the same concept. Fodor's theory also allows for people to possess concepts despite having false beliefs or incomplete information. For instance, one could possess the concept DOG while having false beliefs about dogs or while lacking information about properties that are essential to something's being a dog.[19]

With our sample theory of concepts in hand, we are almost in a position to examine the question of how development proceeds. However, before we can move on to the issue of development, we need to introduce the notion of a *sustaining mechanism*, which turns out to be the key to understanding concept acquisition.[20] A sustaining mechanism is a mechanism in virtue of which a concept stands in the mind–world relation that a causal theory of content, like Fodor's, takes to be constitutive of content. Thus, for Fodor's theory of content, the relevant sustaining mechanisms are ones in virtue of which concepts stand in asymmetric dependence relations with properties in the world. The typical sustaining mechanism of this sort is cognitive or inferential. It helps to determine the semantic properties of concepts, including primitive concepts, but not in the way that the structural elements of a concept with definitional structure determine its content. Rather, a sustaining mechanism determines the semantic properties of a concept indirectly by establishing the mind–world relation that directly determines the concept's content.

Now sustaining mechanisms are likely to vary in different ways from one type of concept to the next (not to mention from one theory of content to the next). Given the extraordinary breadth and diversity of the human conceptual system – ranging from conceptual demonstratives and proper names to the more traditional nominal, verbal, and adjectival concepts – it would be deeply surprising if the same type of sustaining mechanism were at work in each case. Moreover, on a theory like Fodor's, there is no reason to suppose that for any given concept the sustaining mechanisms associated with the concept will even be the same across individuals. Nevertheless, we think it is possible to say something about what some typical sustaining mechanisms might look like for

[19] One might wonder just how uninformed or misinformed someone can be about dogs and still have the concept DOG. According to Fodor's theory of content, someone can be as uninformed or misinformed as you like, as long as she has a representation that stands in the appropriate dependency relations. Another way of looking at the matter is to view the dependency relations of the asymmetric dependence account as implicitly characterizing the range of variability among belief sets that are compatible with concept possession. That is, 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 equivalent to all the others since they all converge on the same mind–world relation. It is this relation, however, and not the specific belief contents, that determine a concept's content. For Fodor, this is a major virtue of the theory. If one tried to explicitly characterize the range of variability that's compatible with a given concept, one would face the daunting task of specifying exactly which beliefs are necessary and which beliefs are sufficient to have the concept. Fodor himself thinks this task isn't merely daunting; he thinks it is impossible. For discussion, see Fodor and Lepore (1992).

[20] This notion was originally introduced in Margolis (1995). The model discussed below is based on the earlier model presented in that work and in Margolis (1998).

different sorts of concepts, and that doing so is both philosophically illuminating and suggestive of a new program of research into the psychological bases of concept acquisition.[21] In illustrating our approach, we will focus on concepts for kinds, and, in particular, concepts for natural kinds.

Perhaps the most important type of sustaining mechanism for a natural kind concept is one that implicates a *kind syndrome* along with a more general disposition to treat instances as members of the category only if they have the same essential property as paradigmatic exemplars of the syndrome.[22] A kind syndrome is a collection of properties that is highly indicative of a kind yet is accessible in perceptual encounters. This may include things like the typical shape, motions, markings, sounds, colors, etc., associated with a kind. The significance of this type of sustaining mechanism is that it readily translates into a learning model. Concept learning – at least for some natural kind concepts, some of the time – can proceed by the accumulation of largely contingent perceptual information about a kind. This information, together with the more general essentialist disposition, establishes an inferential mechanism that can explain why an agent tokens a given concept under the conditions which, according to the asymmetric dependence theory, are constitutive of conceptual content.

In short, we have a sketch of a model of how primitive concepts could be learned and thus the beginnings of a story about how the combinatorial expressive power of a conceptual system could be expanded. To fill out the sketch and to achieve a deeper sense of the issues involved, it pays to turn to experimental work in related areas of psychology. The most promising of these areas is the burgeoning literature on lexical acquisition. This is partly because, on the assumption that language expresses thought, word learning is very closely related to concept learning. Another reason this is a good place to look is that developmental psycholinguists working in this area have discovered a number of important biases that facilitate lexical acquisition by differentially guiding people's reasoning about things of different types – biases that emerge at a very early stage in development and that clearly make contact with people's understanding of natural kinds. We will discuss several relevant findings that help fill out the acquisition model just sketched.

We begin with a result which comes from a series of experiments conducted by Nancy Soja, Susan Carey, and Elizabeth Spelke (Soja, Carey, & Spelke, 1991). Soja et al. were interested in whether young children make an ontological distinction between stuffs (e.g., sand, water, playdoh) and concrete particulars (e.g., pencils, cups, stuffed animals) and

[21] Though Fodor's theory leaves room for the possibility of little or no overlap among the sustaining mechanisms underlying different individuals' possession of the same concept, that doesn't mean that we shouldn't expect an overlap. On the contrary, the sheer speed of acquisition and relative uniformity of conceptual systems across a huge range of concepts provides powerful grounds for thinking there will, as a matter of fact, be considerable uniformity among sustaining mechanisms as well.

[22] Clearly this is not the only possible type of sustaining mechanism for such concepts. One alternative type of sustaining mechanism might involve the possession of a complete true theory of the kind in the extension of the concept, a theory that can act as the final arbiter for processes of categorization. Another alternative might involve the use of experts, deferring to their judgement for resolving issues about the application of a concept. Yet another alternative might involve a sustaining mechanism which is not even cognitive but rather is based on brute psychophysical laws. For discussion, see Margolis (1998).

also whether they do so prior to learning the syntactic cues that mark the difference between mass nouns and count nouns.[23] Adults, of course, think of stuffs differently from how they think of concrete particulars. Stuffs come in quantities of more/less, and you can't count them except in terms of some other mode of individuation. Correspondingly, you can't say there are "two sands"; you have to say there are "two piles of sand". Concrete particulars, on the other hand, come prepackaged in countable units. You don't say "more pencil"; you say "more pencils". Soja et al. discovered that young children also make this ontological distinction between stuffs and concrete particulars, and that they make it prior to learning the syntactic cues that mark it.

Soja et al.'s study was framed in terms of a task where children were expected to learn novel terms for unfamiliar objects and stuffs. In one experiment, children were shown a sample novel object, like a T-junction of brass pipe, which the experimenter referred to using a syntactic construction such as "my blicket" – one that is neutral between objects and stuffs. The children were encouraged to play with the blicket for a while and were then presented with two new things. One was a new T-junction but one clearly made from a different material. The other was a few pieces of piping fragments made from the original material, but this was just an array of bits, so they didn't agree in either number or shape with the original item. The children were then asked to give the experimenter the blicket. The result was that the children, who were 2.5 years old, strongly preferred the choice that agreed with the original item in shape and number, not material. Moreover, when a comparable experiment was performed with a stuff as the target, the children preferred the novel stuff that agreed in material but differed in shape and number. So it seems clear that children at this very young age are subject to a pattern of inference that respects a fundamental ontological distinction. They group things into stuffs and concrete particulars, just as adults do, and reason differentially regarding these things.[24]

Let's now focus on concrete particulars. The Soja et al. experiments suggest that shape is a salient property for young minds. This suggestion is confirmed by other work. Barbara Landau and her colleagues have investigated the role of shape in guiding acquisition of novel count nouns (for a review, see Landau, 1994). In a typical experiment, they introduce a rigid object of a particular shape (e.g., a large wooden "U"-shaped object made with clean right angles), referring to it with count noun syntax (e.g., "See this? This is a dax."). Then they present the child with new objects, ones that vary in either shape, size, or texture, each time asking the child whether it is the same, using the same count noun term and the same revealing syntactic context ("Is this a dax?"). The results are interesting. Three-year-olds accept objects of the same shape 95% of the time, while accepting differently shaped objects only 60% of the time. They clearly show a shape bias, though one that is less pronounced than in adults. What's more, a weaker but still noticeable shape bias is seen among 2-year-olds as well. By the age of 5, children also seem to be able to

[23] Apart from the intrinsic interest of these questions, Soja et al.'s study was meant to test Quine's speculation that children learn the object/stuff distinction by first acquiring the count/mass syntax of their natural language. See Quine (1960).

[24] What's more, their performance doesn't seem to be altered by the presence of informative syntax (e.g., words like "some" and "a"), and their ability to form the stuff/object distinction precedes any facility with these syntactic devices. Soja et al. conclude from these results that, pace Quine, the object/stuff distinction can't be learned on the basis of count/mass syntax. If anything, the learning process goes the other way around.

switch to a texture bias, given the right syntactic context ("This is daxy" or "This is a daxy one"). But the more fundamental ability, the one that occurs earlier, seems to be the ability to generalize on the basis of shape, especially when the presence of a rigid object corresponds with the presence of a novel count noun.

Landau and her colleagues are primarily concerned with the process of lexical acquisition. For them, the central question is how children learn the meanings of their words in the face of the enormous difficulties confronting language learners. A new word could mean just about anything. Even if it is used in front of a cup, say, it could be referring to the color of the cup, the substance it is made of, the texture of its surface, not to mention possible meanings that just happen to be in the spatial-temporal vicinity of the cup, or weird Quinean possibilities – undetached cup parts, cup + table surface, cup time-slice, etc. Contemporary psycholinguists do not tend to view this problem as a mere philosophical curiosity, but rather see it as a challenge to tease out the initial knowledge and biases that make up the standard equipment of the language learner and thereby allow the language learner to weave her way through the myriad possible meanings to the correct one. Landau's work can be seen as providing part of the answer to the question of how a child knows that two items are of the same type. Shape, on this view, is one of the dominant cues that children use. This strategy works because shape provides a defeasible yet highly indicative mark of object kinds, at least at certain levels of a conceptual hierarchy.

Though the shape bias is in the first instance a thesis about language learning, it readily translates into a component of a theory of concept acquisition since the bias clearly constitutes an important part of children's understanding of the nature of objects. The shape bias together with other similar biases and children's implicit understanding of their relative importance enables children to represent kind syndromes. Beginning with shape, children can acquire a concept for some natural kind objects by recording the shape of a novel object and using this in the construction of a sustaining mechanism. The resulting sustaining mechanism, being a syndrome-based sustaining mechanism, will eventually include all sorts of information that is highly indicative of the kind. Yet shape is a good starting point, especially since children have little access to adults' hard-earned knowledge about the vast range of properties that are indicative of different types of kinds.

Still, shape by itself won't do, nor will any combination of simple perceptual features. Such features aren't a perfect guide to kind membership, since, among other things, the world is sometimes populated by what we'll call *fakes* – objects with the same outward appearance of a natural kind which nonetheless are not instances of the category.[25] This is where the essentialist tendency becomes relevant. Putting aside the developmental question for a moment, and just thinking about adults, the idea is that they possess the implicit belief that something is a member of a given natural kind just in case it has the same essential property as paradigmatic exemplars of the kind syndrome. So something may look like a dog, but if it turns out to be an extremely realistic toy (or a large cat on a dark night) instead, then it is no longer deemed a dog, since it does not share the underlying essential property common to all dogs. Similarly, something may fail to look like a dog for whatever reason, but it is still considered to be a dog so long as it has the same underlying

[25] A related problem, though one we'll leave for another time, involves cases of philosophical twins, that is, cases where two distinct kinds have exactly the same outward appearance.

essential property as other dogs. An implicit psychological commitment to essentialism leaves the adult with specific inferential tendencies which serve as a corrective to potential over- or under-generalizations that can be traced to the kind syndrome. The resulting overall inferential tendencies provide a good first-pass model of what one common sort of sustaining mechanism might look like given the asymmetric dependence theory. Psychological essentialism can help to explain why some dependencies are asymmetrically dependent on others. Thus, in the situation where a person sees a fake dog, we basically have a simple case of error. According to the asymmetric dependence theory, though the person might token DOG when seeing the fake, the law that governs this tokening would be asymmetrically dependent on the dog/DOG law. This asymmetric dependence is effectively explained by the person's implicit commitment to essentialism: given psychological essentialism, the fake-dog/DOG law wouldn't hold unless the dog/DOG law did, but not the other way around.

What makes this line of response promising is that ordinary adults do appear to hold a rudimentary form of essentialism. But it is not just adults. Children, too, show signs of an essentialist tendency, one that emerges as young as 2 years old (for a review, see Gelman & Coley, 1991). To give you the flavor of the literature, we will mention just one relevant experiment – our final empirical study in support of the sample model of acquisition. Gelman and Wellman (1991) set out to discover whether children have a grasp of the relevant difference in importance of the insides and outsides of objects. After some preliminary studies that indicate that even 3-year-old children don't necessarily think that similar-looking objects have the same insides, they turned to the larger question of whether the insides or the outsides of various objects are more important in deciding which categories they belong to. In one experiment, they asked 4- and 5-year-olds a series of questions about a range of natural kinds and artifacts depicted by realistic colored drawings. The children were to consider substantial changes to the insides and outsides of the test items and were to report whether the changes affected either an object's identity or its characteristic function. By having the children consider changes to the insides of an object separately from changes to its outsides, the importance of each could be assessed in relation to the other.

The test items fell into two categories: ones which, for adults, the insides are crucial to their identity and functioning (e.g., a dog) and ones for which the insides are irrelevant (e.g., a jar). For each item, the children were asked to consider three transformations. One transformation concerned the insides of the objects. The children were asked to imagine, for example, that the insides of a dog were removed, that the blood and bones and other stuff inside of a dog were taken out, leaving just the outsides, that is, the fur. A comparable transformation concerned the outsides of the objects. For example, the children were asked to consider what would happen if a dog's fur were removed. Also, as a control, the children were asked to consider the situation where the objects moved or were put into different positions or locations. The point of this last transformation was to check whether children have a bias to construe any change in an object as resulting in a change of its identity. In all, then, the children had to answer six questions per test item. For each transformation, they had to report whether the object underwent a change of identity and whether it underwent a change in its ordinary functioning.

The results showed 4- and 5-year-olds to be good at these sorts of questions. The mean

percentage of correct responses for each question type ranged from 65% to 93%. For the insides-relevant items, children were more likely to report that the insides-removal led to a change of identity or function than the outsides-removal, and for the insides-irrelevant items – the containers – they appeared to think that neither the insides nor the outsides were particularly relevant. As Gelman and Wellman (1991) see it, "young children show an impressive ability to penetrate beneath surface appearances" (p. 239). They add in a more speculative tone that "something like an essentialistic disposition could propel knowledge acquisition and shape concept representation early in development – not just at the end" (p. 242). The suggestion that a precocious implicit form of essentialism guides concept learning is very much in accord with the model of acquisition we have been developing here. But what Gelman and Wellman don't say is how acquiring knowledge, even if it is guided by an essentialist tendency, actually results in the acquisition of a new concept. By contrast, the role of psychological essentialism in concept acquisition is clear on the model we have been developing here: psychological essentialism constitutes part of a syndrome-based sustaining mechanism.

Pulling together the various strands of the model we've been developing, we arrive at the following picture of how a new primitive concept could be acquired. The child starts out with a predisposition towards psychological essentialism and a bias to treat shape as especially indicative of kind membership for natural kind objects. She sees a new object that has features that suggest that it is a natural object of some sort. Perhaps it appears to have an internal source of motion, giving it the look of an animate object, or perhaps she hears a novel count noun used of this new item. Either way, upon encountering the item, the child releases a new mental representation and begins accumulating information about the object and linking this to the representation. Giving priority to shape, the child collects and stores a range of information concerning broadly perceptual features of the object. If all goes right, this store comes to embody a kind syndrome; it incorporates information that is highly indicative of the kind and that tends to be exhibited by the kind's paradigmatic instances. Finally, the kind syndrome and the essentialist disposition together govern the inferential tendencies that the child has with respect to the new representation. In particular, they control how she applies the representation to other items and the pattern of corrections she makes, or would make, given further information about why a new item has, or lacks, the syndrome of properties to which she is sensitive. Together these various inferential biases underwrite the dependency relations specified by the asymmetric dependence theory. Later, of course, the storehouse of information that she associates with the representation may grow in all sorts of idiosyncratic ways as she has more interactions with members of the kind. She will continue to have the same concept, however, so long as the information that she associates with the representation establishes the same mind–world dependencies. Again, on the theory of content that's at stake, it doesn't matter what you know about a kind. All that matters is how your concept is hooked up to the world.

Though this picture of concept acquisition is still only the barest sketch of a model of what might be involved in acquiring a new primitive concept, we do think that some significant consequences can be drawn from it vis-à-vis Fodor's Puzzle. We should note again, however, that we do not claim that this is the only model of acquisition. For one thing, the model focuses on just one type of sustaining mechanism, and, for another, it relies on a particular theory of content. All existing theories of content face serious difficulties, and it is

not at all clear which theory is even on the right track. Certainly there is no emerging consensus that, for example, some variant of asymmetric dependence theory is likely to be correct. Along these lines, we suspect that some of the difficulties surrounding our model have less to do with what it says about acquisition than what it says about content; that is, the model inherits all of the problems that infect asymmetric dependence.[26]

On the other hand, we don't think that these issues undermine the present concern. Fodor's Puzzle of Concept Acquisition is solved so long as one can begin to see how someone could come to be in a state that satisfies the conditions of the correct theory of content. As it happens, no one knows what the correct theory of content is. But that shouldn't stop us from developing preliminary models of acquisition and asking the provisional question of how particular theories of content – theories that are on the table – fare when they are scrutinized from the point of view of acquisition. Fodor's own theory of content fares rather well. With just a few psychological principles in place, the theory allows for something rather remarkable – a way of acquiring a new primitive concept, thereby expanding the combinatorial expressive power of the representational system. In short, the picture of the mind driving Fodor's Puzzle is mistaken, since there is a lot of room for acquiring new primitive concepts. Though a complete understanding of how the conceptual system expands is still far off, it is not too early to conceive of possible, even plausible, models for acquiring new primitive concepts.

What's more, the particular model that we have presented is especially suggestive in a number of respects. First, though the model requires a considerable amount of innate structure in the form of biases and inferential mechanisms of various sorts,[27] it still looks like a learning model. That's because it accounts for the acquisition of a concept which, in an important sense, respects the character of one's experience. Seeing a dog doesn't trigger a concept that is already all wired up to go. Rather, seeing a dog initiates a process where information is collected, stored, and manipulated in a way that controls a representation so that it tracks dogs. To our ears, this sounds like learning. Second, the model points towards some interesting and potentially valuable avenues for interdisciplinary research. Philosophers and psychologists often find it difficult to relate to each others' concerns. Philosophers complain that psychological theories of concepts are flawed as accounts of content, while psychologists complain that philosophical theories make no contact with empirical data. But our model shows why philosophers and psychologists may well need one another. What philosophers have to offer psychology is a way of thinking about concepts that ties their nature to the way they are related to the world. What psychologists have to offer philosophy is an empirically viable account of how such mind–world relations are sustained and how they are formed in ontogeny. If this way of thinking is right, then researchers in each of these disciplines can profitably seek guidance and assistance from one another; the results in each field constrain the theoretical options in the other.

[26] For example, one of these concerns whether the theory can distinguish coextensive concepts, including empty concepts. For discussion of this and related problems, see Laurence and Margolis (1999).

[27] Or at least structure that is available prior to learning a concept.

5. Other approaches

We've argued that the key to solving Fodor's Puzzle is the construction of psychological models of concept acquisition in tandem with accounts of the nature of concepts that incorporate a philosophical theory of content. Though the outlines of an answer to Fodor's Puzzle are now clear, the details remain to be worked out. This is to be expected, given the theoretical interdependence of these two areas of research that have until now had very little interaction with one another.

In this section, as a way of shedding further light on Fodor's Puzzle, we'll compare our approach to two others. The first is Fodor's own recent attempt to address the puzzle, one that rejects the psychologically-based strategy that we prefer in favor of a "metaphysical" solution. The second, due to Ned Block, is based on a conceptual role theory of content. We will argue that Fodor's response is deeply unsatisfying in much the same way as the philosophical reactions canvassed earlier in Section 3. Block's strategy, on the other hand, is more in line with our own in that he directly links the issue of acquisition with a theory of content. So we take Block to be an ally. At the same time, though, his response to Fodor's Puzzle faces a distinctive set of challenges.

5.1. Fodor's new metaphysical solution to the puzzle

Recently, Fodor has reassessed his case for radical concept nativism, with the aim of making atomistic theories of concepts more palatable (Fodor, 1998). The pressures that Fodor is responding to are substantial. Atomistic theories, which take lexical concepts to have no internal semantic structure, are widely thought to be nonstarters by psychologists, linguists, and other cognitive scientists. And doubtless one of the major reasons for this reaction is that such theories are thought to be committed to Fodor's radical concept nativism. Levin and Pinker, who we quoted earlier, express widely held beliefs when they say: "psychology… cannot afford to do without a theory of lexical semantics" since "primitive (non-decomposed) mental symbols are the innate ones…" (Levin & Pinker, 1991, p. 4). As should be clear from the foregoing discussion, we believe that these fears are unfounded. But they are real and not unreasonable and they illustrate the substantial influence of Fodor's Puzzle in cognitive science. One can see why Fodor is moved to address them.

Fodor's new view isn't the easiest to understand. The largest part of his discussion is devoted to what he calls the *doorknob/*DOORKNOB *problem*. This is basically the problem of explaining why concept acquisition typically proceeds via causal interaction with things in the extension of a given concept. Why do we typically acquire a concept like DOORKNOB through causal interaction with doorknobs? Fodor's ingenious response focuses on the nature of the properties that our concepts express. He claims that the doorknob/DOORKNOB problem is solved given the principle that most properties are partly constituted by the concepts we employ in recognizing them. For example, under Fodor's analysis something is a doorknob just in case it instantiates the property that human minds "lock on to" given the doorknob stereotype.[28] So it is built into the nature of the property *doorknob* that, given

[28] The reference to "locking" is Fodor's shorthand for the conditions given by an informational theory of content such as the asymmetric dependence theory.

exposure to typical doorknobs (i.e., ones that instantiate the stereotype), people are going to lock on to the property *doorknob*. For this reason, it shouldn't be the least bit surprising that the concept DOORKNOB is acquired by exposure to doorknobs.

How, though, is any of this connected to the issue of concept nativism? As it turns out, Fodor's extended discussion of the doorknob/DOORKNOB problem arises only in connection with a potential objection to his real response to the puzzle. His real response is just to say that concept acquisition may not in fact be explicable in rational/cognitive terms at all. According to Fodor, the moral of his new discussion of concept nativism "may be that though there has to be a story to tell about the structural requirements for acquiring DOORKNOB, intentional vocabulary isn't required to tell it" (Fodor, 1998, p. 143). That is, we don't have any reason to suppose the acquisition story is "in the domain of *cognitive* neuropsychology (as opposed, as it were, to neuropsychology *tout court*)" (Fodor, 1998, p. 143). In effect, Fodor's response calls into question a presupposition of the argument that generates the puzzle of concept acquisition. The presupposition is that concept acquisition is susceptible to psychological explanation. But suppose, instead, that the process is simply a brute neurological process of some sort, as Fodor suggests – one for which there is no corresponding cognitive-level explanation. This would call into question the first premise of the argument for radical concept nativism: concepts might not be acquired by a rational process and yet not be innate either. Fodor's far lengthier discussion of the doorknob/DOORKNOB problem arises only because the doorknob/DOORKNOB problem presents a *prima facie* difficulty for this quick response to the original puzzle. If concepts are acquired by a non-rational mechanism, why is it that they are so often acquired through interaction with items that are in their extension?

Though we find Fodor's response to the doorknob/DOORKNOB problem ingenious, in the end we think that Fodor is being too clever by half. The natural and intuitively compelling solution to the doorknob/DOORKNOB problem is the explanation that our own response to Fodor's Puzzle suggests, namely, a cognitive explanation.[29] While the extent to which the world is mind-dependent is a fascinating and important topic, none of the apparatus that Fodor introduces here is necessary for addressing the issue about innateness.

What's more, Fodor's story about brute neurological mechanisms suffers from much the same problem as some of the other philosophical responses we considered earlier. Though Fodor's story does locate a vulnerable point in the argument for radical concept nativism, it is completely unilluminating. As Fodor has noted elsewhere, "unknown neurological mechanisms" provide no insight into the mental phenomena they are supposed to explain. In the present case, we are left with no explanation of why minds should acquire such things as concepts at all, much less the particular ones that they do acquire, and the acquisition story (such as it is) is left wholly disconnected from any account of the nature of concepts or the psychological processes which operate on them. Thus, Fodor's response to his puzzle amounts to little more than the claim that unstructured concepts may not be innate since it is possible they are acquired by non-psychological mechanisms that no one

[29] The reason why dogs cause the acquisition of DOG, for example, is because dogs are the sorts of things that exhibit the dog syndrome. It is by interacting with dogs that people are able to record this information, store it, and ultimately come to manipulate it, all in a way that establishes the right dependency relations with dogs.

knows anything about. To this, all we can say is that it is a possibility, to be sure, but that granting this isn't saying very much at all.

5.2. Block's conceptual role semantics approach

Ned Block's approach to Fodor's Puzzle is oriented around his commitment to a conceptual role semantics, that is, the view that the content of a concept is determined by its relations to other concepts in a representational system. Block suggests that it is a particular virtue of the conceptual role semantics framework that it can deal with Fodor's Puzzle (Block, 1986, pp. 646–648). We've seen that this isn't true, yet Block's approach is still of interest since the issue of concept acquisition looks somewhat different from the perspective of the theory of content that he favors.[30]

Conceptual role semantics isn't so much a theory of content as a general approach to explaining content. What all conceptual role theories share is the core idea that the content of a concept is to be given in terms of its inferential properties, where these are abstracted from the contributions the concept makes to the inferential properties of the thoughts in which it occurs. For example, the concept for conjunction – AND – occurs in thoughts that exhibit the following inferential patterns:

$$\frac{P \text{ AND } Q}{P} \qquad \frac{P \text{ AND } Q}{Q} \qquad \frac{P,\ Q}{P \text{ AND } Q}$$

A conceptual role semantics, then, might take participating in inferential patterns like these to be constitutive of the concept AND.

How does this basic idea translate into a response to Fodor's Puzzle? There are two aspects to the account. The first is to maintain that concept acquisition is a matter of getting a mental representation to stand in the appropriate inferential relations (the ones that are constitutive of the concept's content). This much follows our general approach to Fodor's Puzzle, according to which theories of concept acquisition should be framed in terms of particular theories of content. It is the second aspect of the conceptual role account – the part that includes a diagnosis of where Fodor's argument goes wrong – that is particularly distinctive of the conceptual role approach. In essence, Block's solution is to deny that there *are* any primitive concepts. According to conceptual role semantics, the content of a given concept is fixed relative to the other concepts in the conceptual system. None need be more primitive or basic than any other.[31] As a result, the primitive/complex distinction can't be used to locate the innate conceptual inventory. From Block's perspective, what Fodor seems to have overlooked is the very possibility that a network of constitutively interrelated concepts can be introduced collectively (Block, 1986, p. 648):

[30] Models of acquisition based on a conceptual role semantics also offer clear advantages for concepts whose contents are unlikely to be captured by any sort of mind–world causal relation – e.g. logical concepts.

[31] The theoretical resources that make this possible are originally due to Frank Ramsey (see Lewis, 1970, 1972 for discussion), but for present purposes the details don't matter.

> One way to see what the [conceptual role semantics] approach comes to is to reflect on how one learned the concepts of elementary physics, or anyway, how I did. When I took my first physics course, I was confronted with quite a bit of new terminology all at once: 'energy', 'momentum', 'acceleration', 'mass', and the like. As should be no surprise to anyone who noted the failure of positivists to define theoretical terms in observation language, I never learned any definitions of these new terms in terms I already knew. Rather, what I learned was how to use the new terminology – I learned certain relations among the new terms themselves (e.g., the relation between force and mass, neither of which can be defined in old terms), some relations between the new terms and old terms, and, most importantly, how to generate the right numbers in answers to questions posed in the new terminology. This is just the sort of story a proponent of [conceptual role semantics] should expect.

What this suggests is that concepts that lack compositional semantic structure may be learned so long as they are acquired by a process that establishes not just their individual inferential roles but also the inferential roles of all of the concepts to which they are constitutively related. So long as all of these inferential roles can be brought about together, all of the implicated concepts can be learned together.

This model contrasts with both traditional accounts and the model that we presented in Section 4. On traditional reductive accounts, new concepts can be acquired one by one, as they are assembled from their constituents. On our model, new concepts can also be acquired one by one, as new sustaining mechanisms are created. In this respect, our model has an important affinity with the reductive tradition that it aims to replace. But Block's model stands apart from both of these approaches in that it doesn't allow for concepts to be acquired one by one; instead, each concept can only be acquired with the simultaneous acquisition of all of the concepts with which it is constitutively interrelated.

Now there are two perspectives from which this suggestion may be evaluated. One concerns whether it provides a cogent response to Fodor's Puzzle by locating a possibility that Fodor's argument overlooks. This we think it does. The other concerns whether the possibility it raises is a promising one. Here things are less clear. This is due, in part, to the fact that conceptual role semantics is less a theory of content than a theoretical approach. Its prospects for explaining concept acquisition turn on the particular type of conceptual role theory envisioned and how the details are spelled out.

Block himself opts for what is called a *two factor* version of conceptual role semantics.[32] On a two factor theory, concepts have two components to their content – one internal, the other external. The external component primarily accounts for the concept's referential properties and is explained in terms of a causal theory. The internal component, in contrast, is supposed to account for the "narrow content" of a concept (i.e., content that may be shared by concepts with different referential properties) and is explained in terms of its conceptual role. The nature of narrow content and whether narrow content is a useful or even coherent notion remain vexed issues in philosophy. For the present purposes, we'll just mention an example that motivates the notion in Block's work. He notes that when

[32] Two factor theories have also been advocated by, amongst others, McGinn (1982), Pinker (1997), and Bloom (2000).

two people utter the word "I", or token the corresponding concept, they refer to different individuals – each to herself – but that there is a semantic property that the two utterances or thoughts share. This semantic property is particularly important to the explanation of behavior. It accounts for why there is a psychological explanation that covers two people who think "I …" even though the two thoughts have very different truth conditions (because they are about different people).

For Block, a complete theory of content can't do without a narrow component, but it can't do without a referential component either. Given his interest in conceptual role semantics, Block, naturally enough, has less to say about the referential component than the narrow component. However, this just underscores how incomplete his account of concept acquisition is. Without a theory of how reference is determined and how mental representations come to acquire their referential properties in development, we've only been given half the story about concept acquisition. So it would seem that Block needs to supplement his conceptual-role-based account of concept acquisition with something like the acquisition model we sketched in the previous section.[33]

Let's put two factor theories aside for the moment, since it is the commitment to inferential role as a determinant of content that is distinctive of the conceptual role approach. One important dimension on which theories adopting this general approach differ is in how big the conceptual roles that determine content are supposed to be. Philosophers who have considered this question have often favored holistic versions of conceptual role semantics, according to which the network that fixes a concept's content includes the concept's relation to nearly every other concept in the conceptual system. One of the main motivations for holism is the fact that it has proven extremely difficult to establish a principled distinction between those inferences that are required for possessing a concept and those that are not. Holists avoid this difficulty by maintaining that the distinction itself is spurious. For holists, you simply can't possess a given concept without possessing all of the other concepts that draw from the same system.

Despite its popularity, holism faces a number of serious challenges. One of these is that that holism renders content exceedingly unstable in the sense that it virtually guarantees that people can't share concepts with the same content (Fodor & Lepore, 1992; Margolis & Laurence, 1998). The reason is quite simple. Different people have different beliefs and perceptions and, consequently, different inferential tendencies. This means that their concepts have different conceptual roles. If content is determined by total conceptual role, then it follows that two different people couldn't have the same concepts. Moreover, the same logic shows that a single individual couldn't possess one and the same concept over time. The problem, of course, is that any given person is constantly updating her perceptions and beliefs and thereby updating the total conceptual roles associated with each and every one of her concepts. As a result, holistic conceptual role theories imply that people undergo massive conceptual change, not just at pivotal moments in childhood, but constantly. And that's not all. Earlier we saw that conceptual role theories address Fodor's Puzzle by insisting that constitutively interrelated concepts are acquired collectively.

[33] Block may in fact welcome this suggestion, since he has expressed sympathy with the view that the referential component of content is determined by the very theory of content on which our model is based (see, e.g., Block, 1993).

However, on a holistic version of conceptual role semantics, what this means is that, to acquire a concept you'd have to simultaneously acquire, not just a few other concepts, but the whole lot of them; you couldn't acquire any one concept without acquiring them all. For these reasons, we think it is safe to say that holistic versions of conceptual role theories will prove to be of little interest to the study of cognitive development.

There are, however, non-holistic versions of conceptual role semantics. Such theories appeal to relatively local networks to fix the contents of particular concepts. These accounts face at least two challenges. One is to show that the holists are wrong and that there *is* a principled distinction between the inferences that are essential to possessing a concept and those that are not. The other is to show for particular concepts that relatively local inferences are strong enough to fix their content uniquely (e.g., closely related concepts like CAT and DOG have to be assigned different local conceptual roles).

At the moment, both of these issues remain unresolved and have generated a great deal of controversy. For example, one of the most widely regarded accounts of how to distinguish constitutive from non-constitutive inferences is Christopher Peacocke's suggestion that the constitutive ones be identified with what he calls the *primitively compelling* inferences. These are supposed to be inferences that, in addition to being compelling, are accepted without being inferred on the basis of other principles and whose correctness requires no further validation – as Peacocke put it, the inferences "aren't answerable to anything else" (Peacocke, 1992, p. 6). Unfortunately, the idea of a primitively compelling inference remains obscure. One wonders, as well, whether it is even suited to the task at hand, since deeply entrenched beliefs would seem to give rise to inferences that are "primitively compelling" yet hardly constitutive of the concepts involved (Rey, 1996). The other challenge is perhaps even more worrying. Most discussions of conceptual role semantics stick very close to the example of logical connectives. However, the concepts for logical connectives are but a tiny and highly idiosyncratic sample of the concepts in our conceptual repertoire. In general, no one knows how to develop a conceptual role account for the vast majority of the rest. But if the theory of content remains at this level of generality – simply amounting to the claim that content is determined by unspecified conceptual roles – then it can offer little direction for theories of concept acquisition.

In closing, we'd like to mention an intriguing possibility: perhaps conceptual role semanticists might allow themselves to be guided by the approach to content and concept acquisition that we argued for earlier. The way to do this would be to reconstrue the inferential dispositions that we cited as part of our syndrome-based sustaining mechanisms, squeezing them into a conceptual role framework. Recall that on our model, these dispositions don't directly determine the content of a natural kind concept like DOG; rather, they set up the mind–world causal relation that does. But what the conceptual role semanticist might try to maintain is that these dispositions do directly determine the content of DOG. In other words, she can claim that what it is to have the concept DOG just is to have the inferential dispositions implicated in the representation of the kind syndrome and the relevant essentialist dispositions. The crucial difference between this sort of account and our own turns on their modal implications. Whereas we appeal to the syndrome-based model as just one important type of sustaining mechanism, the conceptual role account envisioned would claim that the inferential dispositions embodied by the sustaining mechanism are essential to the concepts they support. One clear disadvantage of

doing things this way is that the conceptual role semanticist would face a new version of the problem of achieving stable contents: people who don't possess the conceptual role corresponding to this particular sustaining mechanism wouldn't be able to possess one and the same concept as those who do.[34] Still, the present suggestion gives the conceptual role semanticist a detailed concrete model with many of the virtues of our model.

5.3. Summary

In this section we have looked at two alternative responses to Fodor's Puzzle, one from Fodor and one from Block. Fodor's response is unsatisfying in much the same way as the earlier philosophical responses canvassed in Section 3. It amounts to no more than the claim that concepts might be acquired by non-psychological mechanisms. On the other hand, Block's response is far more interesting and, in certain respects, closer to our own. Block's idea is that conceptual role theories of content allow for essentially new concepts to be learned so long as they are acquired together with all of the concepts with which they are constitutively interrelated. The promise of this approach depends to a considerable extent on how certain outstanding issues within the conceptual role framework are to be addressed. We've noted some of the peculiar challenges that conceptual role theorists face; it remains to be seen whether these can be overcome.

For the present purposes, however, the crucial point we want to emphasize is our broad agreement with Block that questions about the nature of concepts are intimately bound up with questions about how they are acquired. If one considers the precedents in cognitive science, it seems only natural to suppose that these issues should be inextricably linked. Perhaps *the* driving motivation behind the voluminous and highly productive work in generative grammar has been the desire to provide an account of language that does justice to how it is acquired. In the study of language, this orientation has paid off tremendously. Why should it be any different with the study of the conceptual system? In fact, if one looks at the larger tradition of western philosophy, the idea hardly seems new. Philosophers have often seen questions about acquisition as being tightly connected to questions about the nature of conceptual content. It is just that they have usually assumed that concepts have definitional structure and thus have also assumed that acquisition respects this structural constraint. What's new to our suggestion is that, even in the absence of definitions, Fodor's Puzzle presents no genuine barrier to accounting for how a concept is learned. So even with primitive concepts, an investigation into how they are acquired seems likely to say quite a lot about their nature.

[34] The conceptual role semanticist might try to get around this problem as well by loosening her account so that the identity conditions for a concept's content aren't framed in terms of a particular conceptual role but rather in terms of a set of conceptual roles that are taken to be equivalent for the purposes of content determination. This would allow her theory to mimic our own account in even more detail. She could then say that *any* set of inferential dispositions that support the appropriate mind–world relation should be included in the set. At this point, however, it would be difficult to see why the conceptual role semanticist shouldn't simply endorse the casual theory of content that is clearly guiding her.

6. Concluding remarks

Fodor's Puzzle is important both because it continues to exert pressure on philosophers and cognitive scientists (whether they recognize this influence or not) and because it is connected to deep issues about the nature of concepts. We have argued that the correct response to Fodor's Puzzle involves a fundamental reorientation in thinking about concepts yet one that is independently compelling. Rather than constructing theories of content in isolation, philosophers and psychologist have to pool their resources and develop theories of content in tandem with accounts of how people could come to be in states that satisfy the conditions that these theories impose. We presented our strategy in a schematic form to show that it isn't tied to a single model but rather to a family of models, all of which employ the same general means for answering Fodor's Puzzle. We went on to present a concrete illustration of the general approach using Fodor's own theory of content. Needless to say there is much work to be done in filling out the model. Yet by showing that, in principle, the combinatorial expressive power of the conceptual system can be expanded, we take ourselves to have provided a thorough response to Fodor's Puzzle.

Does our rejection of radical concept nativism amount to a vindication of empiricism? One might think it does on the grounds that our detailed model of acquisition has elements of a learning model and that learning is at the heart of empiricism. But the model we appeal to in response to Fodor's radical concept nativism is hardly one that a true empiricist would want to endorse. Empiricist models have relatively little innate structure as a precursor to concept acquisition. What's more, the structure that one finds in empiricist models tends to be very domain general in its application, that is, the very same mechanism that is suited to the acquisition of one concept is supposed to be suited to the acquisition of most others.[35] But our model suggests that the cognitive resources that are involved in concept acquisition are quite rich. And there's no reason to think that these have general application. Our own hunch is that there are a variety of resources that are suited to the development of different types of sustaining mechanisms. In each case, what you'd have is a relatively (but not wholly) general disposition to acquire a sustaining mechanism of a certain type. Later, given the character of experience, particular sustaining mechanisms and hence particular concepts would be the natural outcome.

It is easy to underestimate the resources required for processes of cognitive development. Generations of empiricist theories bear witness to just how difficult it is to account for the learning of concepts that we all take for granted. For instance, take Locke's treatment of the concept of LYING (in the sense of telling falsehoods). After providing a sketch of an analysis that includes reference to minds, speakers, and "signs put together by affirmation or negation, otherwise than the *Ideas* they stand for", he adds that it should be

[35] We take it that these are the two most important features of empiricist thought insofar as the empiricist/nativist debate has contemporary significance. Everyone believes there is a certain amount of innate structure. The question is how much. Cf. Chomsky (1975, p. 13): "Every 'theory of learning' that is even worth considering incorporates an innateness hypothesis. Thus, Hume's theory proposes specific innate structures of mind and seeks to account for all of human knowledge on the basis of these structures. The question is not whether learning presupposes innate structure – of course it does; that has never been in doubt – but rather what these innate structures are in particular domains".

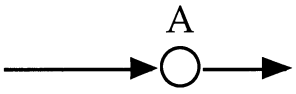

Fig. 1. The image schema for caused motion (adapted from Mandler, 1994).

clear that the final analysis of the concept will involve primitive, sensory concepts. "[I]t 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" (Locke, 1690/1975, p. 292). Unfortunately, it is all but obvious how the analysis is supposed to go. Concepts like SPEAKER and MIND are not themselves sensory, and it is far from clear (to say the least) how they might be reduced to other concepts that are patently sensory.

To take another example – one that isn't merely of historical interest – consider Jean Mandler's influential discussion of concept acquisition in her paper "How to build a baby, II" (Mandler, 1994). Mandler's aim is to show how babies can learn fundamental concepts such as CAUSE, AGENT, and SUPPORT, all short of their possessing a discursive system of representation and, in particular, a language of thought. Her suggestion is that infants come into the world with a disposition to form *image schemas*, which are supposed to be non-discursive representations that encode spatial-temporal information on the basis of perceptual input. For example, Mandler gives the representation depicted in Fig. 1 for CAUSED MOTION, which concerns the sort of motion that results when one object collides into another. Now it is important not to over-interpret the representation. In particular, you aren't allowed to interpret the object with the letter above it as something being acted upon (at least not yet). At most, what's represented is that one item moves next to another and that, when they are right next to one another, the second item starts to move. Somehow from this early quasi-perceptual representation the concept CAUSED MOTION is supposed to emerge.

But there are at least two problems with this proposal, both of which show that, far from being a novel theory, Mandler's account is surprisingly recidivistic. One is that she relies upon a resemblance theory of content, the view that a representation represents what it does by virtue of resembling it. Though we won't rehearse the problems with this view, we take it that a resemblance theory simply can't be made to work (see Fodor, 1975; Wittgenstein, 1953). The other is that she has nothing to say concerning how the spatial-temporal representation – her image schema – forms the basis of the concept CAUSED MOTION. Notice that the concept CAUSED MOTION, because it involves the concept CAUSE, is of interest precisely because it far outstrips the spatial-temporal properties that it may contingently track. To view a scene as a causal event is to view one object as *acting upon* another. The two objects must be assigned distinct causal roles, not just distinct spatial-temporal positions. Since adults know this, and arguably even infants know this (see Leslie, 1982), Mandler's image schemas fail to reconstruct the concept that she takes to be her target. The interest of this example is that it shows just how hard it is to work out a theory of concept learning. Supposing Mandler is right that image schemas play an important role in the acquisition of concepts, it is still a great mystery how one is supposed to get from the relatively impoverished content of a representation like the one given in

Fig. 1 to the far richer content that goes with the concept CAUSED MOTION. Without any insight on this matter, we have no reason to think that the latter is even learned. For all we know, it is an innate concept that's triggered by certain spatial-temporal patterns.

In the end, we wouldn't be bothered in the least if empiricists could find a way to show that the inferential tendencies we cite in our own model are learned on the basis of more general cognitive capacities. But at the same time, we see no reason to think that they will. As with the case of language, the issue shouldn't be one of deciding in advance what maximal amount of innate cognitive machinery is palatable. Rather, the issue is whether our theories incorporate a rich enough collection of innate capacities, processes, representations, biases, and connections to accommodate the basic facts of cognitive development. In any event, our primary concern here is with Fodor's Puzzle, not the larger debate between empiricists and nativists. And with Fodor's Puzzle, the situation should now be clear. The model that we have presented shows that Fodor's radical concept nativism can be avoided and that it is possible for the combinatorial expressive power of the human conceptual system to be expanded. Contrary to the view many in cognitive science share with Fodor, primitive concepts can be learned.

Acknowledgements

This paper has been in the works for longer than either of us cares to remember. We've benefited from many conversations with friends and colleagues and are grateful for their comments. We'd especially like to thank Jerry Fodor, Barbara Landau, Alan Leslie, Kenneth Taylor, Timothy Williamson, and audiences at the University of Warwick, Hampshire College, the University of Edinburgh, the University of Hull, the University of Sheffield, University College London, University of Wales Lampeter, University of Nottingham, Rice University, Washington University St. Louis, and the 1999 "Philosophy, Mind, and Society Conference" held in Turin, Italy. We'd also like to thank the two anonymous referees. Finally, E.M. would like to thank Rice University's Center for the Study of Cultures for supporting this project.

References

Block, N. (1986). Advertisement for a semantics for psychology. In P. A. French, T. E. Uehling Jr. & H. K. Wettstein (Eds.), *Midwest studies in philosophy, X: Studies in the philosophy of mind*. Minneapolis, MN: University of Minnesota Press.

Block, N. (1993). Holism, hyper-analyticity and hyper-compositionality. *Mind and Language*, *8* (1), 1–26.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Chomsky, N. (1975). *Reflections on language*. New York: Pantheon.

Chomsky, N. (1991). Linguistics and cognitive science: problems and mysteries. In A. Kasher (Ed.), *The Chomskyan turn*. Oxford: Blackwell.

Churchland, P. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Clark, A. (1994). The language of thought. In S. Guttenplan (Ed.), *A companion to the philosophy of mind*. Oxford: Blackwell.

Cowie, F. (1999). *What's within: nativism reconsidered*. Oxford: Oxford University Press.

Di Sciullo, A., & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

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.

Fodor, J. A. (1990a). A theory of content, I: the problem. In *A theory of content and other essays*. Cambridge, MA: MIT Press.

Fodor, J. A. (1990b). A theory of content, II: the theory. In *A theory of content and other essays*. Cambridge, MA: MIT Press.

Fodor, J. A. (1998). *Concepts: where cognitive science went wrong*. Oxford: Oxford University Press.

Fodor, J. A., Garrett, M., Walker, E., & Parkes, C. (1980). Against definitions. *Cognition*, *8*, 263–367.

Fodor, J. A., & Lepore, E. (1992). *Holism: a shopper's guide*. Oxford: Blackwell.

Fodor, J. D., Fodor, J. A., & Garrett, M. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, *6*, 515–532.

Gelman, S., & Coley, J. D. (1991). Language and categorization: the acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: interrelations in development*. Cambridge: Cambridge University Press.

Gelman, S., & Wellman, H. (1991). Insides and essences: early understandings of the non-obvious. *Cognition*, *38*, 213–244.

Goodman, N. (1954). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Hume, D. (1978). *A treatise on human nature*. Oxford: Oxford University Press. (Original work published 1739)

Jackendoff, R. (1989). What is a concept, that a person may grasp it? *Mind and Language*, *4*, 68–102. (Reprinted in Jackendoff, R. (1992). *Languages of the mind: essays on mental representation*. Cambridge, MA: MIT Press)

Kripke, S. (1982). *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.

Landau, B. (1994). Object shape, object name, and object kind: representation and development. In D. Medin (Ed.), *The psychology of learning and motivation*. New York: Academic Press.

Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: core readings*. Cambridge, MA: MIT Press.

Leslie, A. (1982). The perception of causality in infants. *Perception*, *11*, 173–186.

Levin, B., & Pinker, S. (1991). Introduction. In B. Levin & S. Pinker (Eds.), *Lexical & conceptual semantics*. Oxford: 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. (1975). *An essay concerning human understanding*. Oxford: Oxford University Press. (Original work published 1690)

Mandler, J. (1994). How to build a baby, II: conceptual primitives. *Psychological Review*, *99*, 587–604.

Margolis, E. (1995). *Concepts and the innate mind*. Unpublished PhD dissertation, Rutgers University, New Brunswick, NJ.

Margolis, E. (1998). How to acquire a concept. *Mind and Language*, *13*, 347–369.

Margolis, E., & Laurence, L. (1998). Multiple meanings and the stability of content. *Journal of Philosophy*, *95* (5), 255–263.

McGinn, C. (1982). The structure of content. In A. Woodfield (Ed.), *Thought and object: essays on intentionality*. Oxford: Oxford University Press.

Miller, G. (1978). Semantic relations among words. In M. Halle, J. Bresnan & G. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.

Pinker, S. (1994). *The language instinct*. New York: William Morrow.

Pinker, S. (1997). *How the mind works*. New York: W.W. Norton.

Putnam, H. (1988). *Representation and reality*. Cambridge, MA: MIT Press.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Rey, G. (1996). Resisting primitive compulsions. *Philosophy and Phenomenological Research*, *LVI*, 419–424.

Samet, J. (1986). Troubles with Fodor's nativism. In P. A. French, T. E. Uehling Jr. & H. K. Wettstein (Eds.), *Midwest studies in philosophy, X: Studies in the philosophy of mind*. Minneapolis, MN: University of Minnesota Press.

Samet, J., & Flanagan, O. (1989). Innate representations. In S. Silvers (Ed.), *Rerepresentation*. New York: Kluwer Academic.

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

Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions on word meaning: object terms and substance terms. *Cognition*, *38*, 179–211.

Sterelny, K. (1989). Fodor's nativism. *Philosophical Studies*, *55*, 119–141.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.