Margolis, E., & Laurence, S. (2011). Learning Matters: The Role of Learning in Concept Acquisition. *Mind & Language*, 26:5, pp. 507-639.

DOI: 10.1111/j.1468-0017.2011.01429.x

The final publication is available at the Wiley Online Library:
https://onlinelibrary.wiley.com/doi/10.1111/j.1468-0017.2011.01429.x

Learning Matters: The Role of Learning in Concept Acquisition

Eric Margolis & Stephen Laurence*

Abstract: In *LOT 2: The Language of Thought Revisited*, Jerry Fodor argues that concept learning of any kind—even for complex concepts—is simply impossible. In order to avoid the conclusion that all concepts, primitive and complex, are innate, he argues that concept acquisition depends on purely biological processes. In this paper, we show (1) that Fodor fails to establish that concept learning is impossible, (2) that his own biological account of concept acquisition is unworkable, and (3) that there are in fact many promising general models for explaining how concepts are learned.

1. Introduction

Where do our concepts come from? For most philosophers and psychologists, the short answer is that the vast majority of human concepts are learned. If it weren't for concept learning, it would be hard to see how science, technology, and culture could ever have developed. There would be no microscopes, no vaccines, no television. Forget electronic books; we wouldn't even have paper. But while it may seem obvious that concept learning is a fact of life, the question of how even fairly ordinary concepts can be learned can quickly lead to perplexity. Often learning a new concept involves creating new representational resources that exceed those you possessed prior to learning the new concept. The problem is to explain how an essentially richer system of representation can be formed on the basis of a more impoverished one.

* This paper was fully collaborative; the order of the authors' names is arbitrary. EM would like to thank Canada's Social Sciences and Humanities Research Council for supporting this research.

No one has done more to highlight this difficulty than Jerry Fodor. Fodor has been a persistent and outspoken critic of theories of concept learning—drawing conclusions most have found outrageous—ever since he first broached the issue in *The Language of Thought*. It is now more than thirty years later, and with the publication of *LOT2: The Language of Thought Revisited*, Fodor has concluded, of all things, that his previous discussions showed 'a failure of nerve' for not going far enough (2008, p. 138).[1] In the past, Fodor's argument against learning was based on broadly empirical considerations and included the loophole that complex concepts might be learned. But in LOT2, Fodor claims to have an a priori argument against concept learning and has eliminated his earlier qualification regarding complex concepts. His current view is unapologetically and categorically opposed to learning. 'What I should have said is that it's true and a priori that the whole notion of concept learning is per se confused. *Punkt*' (2008, p. 130). There is also a second important change of heart in LOT2. In earlier work, Fodor had taken nativism to be the alternative to concept learning, leading to his trademark view that virtually all so-called *lexical concepts*—roughly, concepts corresponding to single natural language words—are innate. LOT2 abandons this radical nativism, coupling the revised argument against learning with the view that there may be no innate concepts at all. According to LOT2, the way to avoid innate concepts is to stop searching for a theory of acquisition that is framed in terms of intentional states and processes. Instead, we should embrace an account of concept acquisition that depends on noncognitive and nonrational—or as Fodor puts it, *brute*—biological processes.

In this paper, we critically examine Fodor's current thinking about concept acquisition, both his argument against learning and his proposal that concept acquisition is best seen as a biological rather than a cognitive phenomenon. Though we'll argue that Fodor is wrong on both counts, we believe that much can be learned from seeing where he goes wrong. One of the advantages of Fodor's new stripped-down argument against concept learning is that it greatly clarifies the motivations that have been driving his skeptical position from the beginning. But even more important is that a proper treatment of the argument highlights a number of ways in which concepts may be learned that have been overlooked

[1] We will use the abbreviations LOT1 and LOT2 (without italics) to refer to Fodor (1975) and Fodor (2008), respectively.

not just by Fodor but by his critics as well. We identify and sketch a number of these ways in an effort to explicate the true power of learning. The fact is that learning really does matter, though how learning is able to achieve what it does is rightly controversial and still only dimly understood.

2. Fodor's New Argument Against Concept Learning

In this section, we provide an overview of Fodor's new argument against concept learning and how it emerged as a reaction to his earlier discussions.

Fodor's skepticism about concept learning has been a lasting theme in his work. After LOT1, Fodor revisited the issue a number of times and came to emphasize that the question of whether concepts can be learned turns, to a large extent, on whether they have internal compositional structure. What he imagined was that a concept like BIG RED BARN is at least a candidate for being learned because it can be assembled from its constituents, but that a concept like RED (assuming it has no compositional structure) can only be activated by an innate triggering mechanism.[2] In Fodor (1981), Fodor was so focused on the issue of conceptual structure, that the bulk of his case for his notorious *radical concept nativism*—the view that virtually all lexical concepts are innate—consisted in presenting evidence against lexical conceptual structure, that is, evidence that word-sized concepts are not composed out of simpler concepts. His nativist thesis never won over many adherents, but in the larger cognitive science community, the proposed link between learning and conceptual structure was well-received. There was a sense that Fodor was right about identifying the innate concepts with the simple or primitive (that is, unstructured) ones and that learning, of necessity, involves constructing new complex concepts from a stock of innate primitives. If few scientists went on to embrace Fodor's radical concept nativism, this was because they disagreed with Fodor about the prospects for identifying the needed conceptual structure. The feeling was that Fodor had the overarching dialectic right but that he was wrong about the empirical issues regarding conceptual structure.

Fodor hasn't wavered on the question of whether lexical concepts have structure. He continues to maintain that ordinary lexical concepts (GIRAFFE, FORK, RUN, etc.) are

[2] We follow the convention of using expressions in smallcaps to refer to concepts.

primitives. But a major change in LOT2 is that he no longer thinks that the issue of conceptual structure matters to the question of whether concepts can be learned. 'I must confess that I have come to agree with my critics that there is something wrong with the argument as *LOT1* presented it; namely, that the conclusion is too weak and the offending empirical assumption—that quotidian concepts are mostly primitive—is superfluous' (2008, p. 130). Fodor's new argument is thus intended to be a far more powerful version of the earlier argument. It draws much the same anti-learning conclusion but without the need to take a stand on whether any given concept has structure or not. The result is that all concepts are treated the same. On Fodor's current view, BIG RED BARN is as impossible to learn as RED. The whole idea of concept learning is simply confused.

Here is our reconstruction of the LOT2 argument:

1. Concepts (whether primitive or complex) cannot be learned via hypothesis testing.
2. There is no other way that a concept could be learned.
3. Therefore, concepts can't be learned.

Again, we should emphasize that Fodor's conclusion stops short of claiming that unlearned concepts are innate. But the conclusion isn't any more agreeable. Fodor still maintains the bewildering view that no concept is ever learned.

In support of the first premise, Fodor argues that if concepts were acquired via hypothesis testing (HF)[3], the process would be inherently circular (2008, p. 139):

> Now, according to HF, the process by which one learns C must include the inductive evaluation of some such hypothesis as 'The C things are the ones that are green or triangular'. But the inductive evaluation of that hypothesis itself requires (*inter alia*) bringing the property *green or triangular* before the mind as such. … Quite generally, you can't represent anything as *such and such* unless you already have the concept *such and such*. All that being so, it follows, on pain of circularity, that 'concept learning' as HF understands it *can't* be a way of acquiring concept C. … Conclusion: *If concept learning is as HF understands it, there can be no such thing*. This conclusion is entirely general; it doesn't matter whether the target concept is primitive (like GREEN) or complex (like GREEN OR TRIANGULAR).

---

[3] HF is Fodor's shorthand for the view that 'concept learning is a process of inductive inference; in particular, that it's a process of projecting and confirming hypotheses about what the things that the concept applies to have in common' (2008, p. 132; italics removed).

In other words, to test and confirm a hypothesis concerning the identity conditions for a given concept, you must already be able to entertain the concept. But if you can already entertain the concept prior to learning it, then you aren't really learning it; you already have it. So no concept, not even a complex concept, can be learned in this way.

Of course, the problem with hypothesis testing models wouldn't be so bad if other approaches to concept learning were viable. The burden of Fodor's second premise, however, is to exclude all other approaches in one fell swoop. Fodor motivates hypothesis testing models by noting the intuitive contrast between learning and instances where a concept is acquired through wholly nonrational processes (2008, p. 135):

> [T]he experience from which a concept is learned must provide (inductive) evidence about what the concept applies to. Perhaps COW is learned from experiences with cows? If so, then experiences with cows must somehow witness that it's cows that COW applies to. This internal connection between concept learning and epistemic notions like evidence is the source of the strong intuition that concept learning is some sort of rational process. It contrasts sharply with kinds of concept acquisition where, for example, a concept is acquired by surgical implantation; or by swallowing a pill; or by hitting one's head against a hard surface, etc.

Hypothesis testing comes into the picture because, as Fodor sees it, there aren't any alternative accounts that can do justice to this contrast. Hypothesis testing is the only proposal in which concept acquisition is rationally constrained. '[I]f we are given the assumption that concept learning is some sort of cognitive process, HF is de facto the only candidate account of what process it might be' (2008, p. 139). Granted, there may be nonrational processes that eventuate in a concept's being acquired, but *learning* can only unfold as the formulation and testing of hypotheses.

In sum, though we need hypothesis testing for a concept's acquisition to count as learning, hypothesis testing is itself a nonstarter as an explanation for concept learning since it presupposes that the concepts to be learned are already possessed. The upshot, as Fodor puts it, is that 'there can't be any such thing as learning a concept' (2008, p. 139).

3. Learning Complex Concepts

Though Fodor's new argument is indifferent to whether a concept is primitive or complex, we think it helps to look at these two cases separately, since they raise different issues.

We'll start with complex concepts in this section and will postpone the discussion of primitive concepts until section 4.

Now the issue about complex concepts itself breaks down into two related questions. First, is it true that complex concepts can't be learned by hypothesis testing? Second, is it true that learning a complex concept requires hypothesis testing? LOT2 gives a firm 'yes' to both of these questions. We argue, on the contrary, that hypothesis testing models for complex concepts are perfectly viable and consequently that complex concepts can be learned according to Fodor's preferred understanding of what learning requires. But, in addition, complex concepts can be learned in other ways as well, and so proponents of learning shouldn't feel restricted to the hypothesis testing framework. If our analysis is right, then, pace Fodor, the prospects for complex concepts are rather promising. There are more than enough resources to explain how they can be learned.

3.1. Hypothesis Testing Defended: Case 1

Fodor's case that complex concepts cannot be learned via hypothesis testing turns on his claim that hypothesis testing models are circular in that they presuppose the very concepts whose learning they are supposed to explain. Fodor's discussion of this claim in LOT2 is organized around the example of someone trying to learn the concept GREEN OR TRIANGULAR. This example is unfortunate in that GREEN OR TRIANGULAR is a somewhat unnatural concept whose acquisition occurs in an unspecified context. Among other things, the artificial nature of Fodor's example obscures the fact that his discussion focuses on a single moment in the hypothesis testing process—the instant at which the correct concept is formulated—and overlooks the significance of the stages that precede it. With a more realistic example in hand, it pays to examine the entire process as it unfolds, starting with the learner's initial hypotheses.

Consider a fairly ordinary example of what certainly looks like concept learning. An intermediate-level dancer undertakes to learn a new dance by representing its component moves to herself. One strategy that the dancer might adopt, though by no means the only one, is to master a complex concept that describes the sequence of moves. Suppose our learner is enrolled in a course by the Royal Scottish Country Dance Society but happens to miss the class that covers Maxwell's Rant, a dance that involves the following sequence:

reflection reels of three on opposite side, followed by reflection reels of three on own side, followed by crossing with right hands, followed by casting off, followed by a half figure of eight, followed by leading down the set, followed by casting up, followed by turning with right hands. We can imagine that our learner watches from the sidelines while her classmates practice the routine the next time they meet. The dance is complicated. Not surprisingly, the learner's first attempt at representing the sequence isn't quite right and she is well aware of having made some mistakes. This leads her to watch the dance again, to make some corrections, and then repeat the process, making further corrections and filling in the gaps in her representation of the sequence. Gradually she builds a more thorough and accurate representation and eventually is down to a single omission. So she watches one last time and says to herself 'Of course! Before the final turning with right hands, I need to cast up'. At this point, she still hasn't explicitly entertained the final representation of the dance, but the realization that she has only missed this one step causes her to confidently formulate and adopt the complex concept that captures the full sequence of moves: REFLECTION REELS OF THREE ON OPPOSITE SIDE, FOLLOWED BY REFLECTION REELS OF THREE ON OWN SIDE, FOLLOWED BY CROSSING WITH RIGHT HANDS, FOLLOWED BY CASTING OFF, FOLLOWED BY A HALF FIGURE OF EIGHT, FOLLOWED BY LEADING DOWN THE SET, FOLLOWED BY CASTING UP, FOLLOWED BY TURNING WITH RIGHT HANDS. This sort of example highlights the fact that hypothesis testing involves a *process* that unfolds through a series of steps, and does not merely consist of the single moment at which the correct concept is formulated.

The key thing to notice about this example is that it is completely immune to Fodor's charge of circularity. This is because all of the confirmation of the relevant hypothesis takes place *before* the final concept is even explicitly formulated. The learner, you will recall, is completely confident about the concept she is to learn as soon as she notices the final move that had eluded her and *before* she explicitly entertains the full concept. After the hard work of formulating and rejecting her earlier hypotheses, she knows exactly what the identity conditions are for the target concept, and it only remains for her to construct the concept and be done with it. But if all of the justification occurs prior to constructing the concept, then the concept doesn't have to be in place before it is learned. On the contrary, it is because of the justification that came before its appearance that the concept even enters the learner's mind.[4]

What this example shows is that it is perfectly possible to learn a new complex concept through a hypothesis testing procedure. There is absolutely no threat of circularity in the account because the learned concept appears on the scene after the justification occurs.

Of course, there are other possible ways in which the dancer might have come to possess the concept and where it might make sense to say that although the concept is acquired, it isn't really learned. For example, the concept could have been innately specified, not in the trivial sense of being a possible arrangement of innate representational building blocks, but in the highly substantive sense of being fully encoded as a prefabricated complex concept. Or the learner could have formulated the correct concept simply as a lucky wild guess, completely independent of the evidence, before she even saw the dance being performed. Neither of these situations is especially plausible for a complex concept like the one in our example, but that isn't the point. What matters is just that, had they occurred, there would be little reason to say that the concept was learned. But in the example we are considering, the concept has a very different kind of origin. It is assembled as the need arises, and the process that is responsible for its occurrence, as well as its persistence in the agent's mind, is one that is sensitive to the agent's observations and her previous attempts to accommodate them.

Finally, it is worth adding that there is nothing particularly special about the concept that is learned in the example. The same considerations apply, at the very least, to any complex concept that describes a sequence of events in terms of a more basic stock of event types (e.g., concepts involved in learning a chess strategy, a cooking recipe, a new type of knot, or a chord change). In LOT2, Fodor characterizes his argument against learning as an a priori argument and claims to have located a confusion that is inherent to the hypothesis testing framework. But the example we have given, and the range of cases that it illustrates, shows that Fodor's circularity objection is misplaced. Hypothesis testing

[4] The hypothesis confirmation notably also does not proceed via enumerative induction. But then many of the most interesting hypotheses in science and everyday life aren't confirmed by simple enumerative induction. Fodor's focus on enumerative induction is another way in which his case is illicitly biased to make the circularity claim seem more plausible.

models remain perfectly viable for complex concepts so long as we take into account the way that the learning process actually plays out.

3.2. Hypothesis Testing Defended: Case 2

In the case we have just considered, there isn't even the appearance of circularity given that the explicit representation of the concept's identity conditions occurs after the justification has taken place. But many instances of concept learning will not be so clean-cut.[5] For example, consider the simple modification to the dance example where the learner isn't so confident about the final correction to her representation of the dance sequence. Then she might formulate the hypothesis that captures the full and complete dance but seek further evidence regarding its accuracy. If that is how it goes, then it would appear that she *does* have to entertain the concept to be learned in advance of executing all of the inferences that figure in the learning process. Now one possibility is that, in this type of case, concepts cannot be learned via hypothesis testing; concept learning via hypothesis testing might be restricted to the original type of case we considered (Case 1). No doubt, Fodor would insist that in the modification we are envisioning, which is arguably more typical, hypothesis testing models are circular. But we think that the threat of circularity isn't a serious worry even here. We will discuss three responses to the circularity charge regarding the modified case of concept learning via hypothesis testing. Each response offers a different way to make sense of how concepts could be learned. Though we think that all three types of response are defensible, only one is needed in order to establish that complex concept learning via hypothesis testing is possible in such cases and therefore that concept learning via hypothesis testing is extremely common.

Response 1. Let's consider the structure of the modified version of the dance example in a bit more detail. Unlike in the original case, the learner is unsure of the final correction and feels the need to confirm the hypothesis that incorporates the concept whose learning is in question. The course of events in this modified example can be broken down into three stages. (1) As the learner observes her fellow students, she tries to capture the dance's

[5] To undermine Fodor's claim that it is true a priori that complex concepts cannot be learned via hypothesis testing, only a single counterexample is required. Nonetheless, it is instructive to explore a range of cases for the issues they raise.

essential structure, formulating several complex concepts that represent the hypothesized sequences. (2) She arrives at a complex concept that captures the full structure of the dance but she is still unsure about whether she has it right. (3) She goes on to confirm her hypothesis through further viewings of the dance. Now Fodor assumes that the learning doesn't happen until stage (3) is complete. But there is a plausible case to be made that the concept is already learned by stage (2). Seeing this requires breaking the concept learning process down even further.

We can distinguish between an agent first learning a new concept (learning that there is a concept of type C) and subsequently learning an important fact about the concept (learning that C accurately describes the dance). If we analyze things in this way, then the learner doesn't just learn the concept C. She learns a series of concepts, namely, the series of concepts she explicitly formulates in watching the dance. These concepts will each describe dance sequences related to the sequence involved in the observed dance. Some, which omit moves in the observed dance, will describe shorter dances. Others, where the ordering of moves has been misrepresented, will describe dances that involve permutations of moves in the dance sequence. But all of these concepts will represent sequences of dance moves that the learner is likely not to have previously entertained. They will also derive from causal interaction with the environment in which the agent sets out to learn a new concept, and where the resulting concept is composed of component representations based on the learner's assiduous observations. It is not implausible to suppose that this type of causal and informational environmental sensitivity amounts to a rational sensitivity sufficient to count as learning a new concept—it is certainly very different than acquiring a new concept by surgical implantation or by miraculously being hit on the head in just the right way. If the concept is already learned at stage (2), what happens at stage (3)? On this way of looking at things, something is still learned at stage (3), but not the concept. Rather, what the learner learns at stage (3) is an important fact regarding the (previously learned) concept, namely, that it does justice to the observed dance's structure. Stage (3), in other words, is not necessary for concept learning, though it may well play a key role in determining whether the learned concept is retained as a stable part of the agent's conceptual repertoire.

By theoretically separating two closely related things that are learned in typical concept learning situations— learning *that there is a concept C* and learning *that C is the right concept* —we can see how a concept could already have been learned at stage (2), even though the learner remains unsure as to whether this concept accurately describes the dance she has observed. This response gives us a way to understand how a new complex concept could be learned via hypothesis testing even before all the justificatory processes involving that concept have been completed. And since this response has the concept learned prior to stage (3), there is again no circularity involved in its acquisition.

Response 2. The key insight of the first response is that a certain sort of sensitivity to the environment in the construction of a new concept may suffice for the concept to be learned. However, there is still something to be said for the view that the justificatory process in stage (3) is important enough that we should deem it part of what goes into learning the concept. One reason to maintain this higher standard is the recognition that concept learning is often a directed process. Given that the agent has the aim of learning the concept in question, confirming that it is the concept she aimed to learn is a necessary part of the concept learning process; simply having reasons for entertaining the concept isn't enough.

Suppose we take this perspective and assume that the later confirmation at stage (3) is part of what goes into learning the concept. Does this entail that the concept is not learned until after stage (3) is completed and thus that concept learning is inherently circular? No. Even if we accept that stage (3) is essential to learning the concept, and that stage (3) takes place after the acquisition of the concept at stage (2), we still aren't required to say that the learning takes place after the concept is acquired. To see why, it helps to consider an analogous issue from a different domain. In the present case, a concept is acquired at time t_1, the confirmation that the concept accurately describes its target is essential to this being a case of concept learning, and this confirmation happens at a later time, t_2. Now consider what Davidson (1969) says about the time of a killing. Davidson describes a case where he pours poison into the water supply at time t_1 with the intention of killing a traveler and that his action has precisely this effect at some later time, t_2. When does the killing take place? As Davidson notes, 'the most usual answer is that my killing the traveler is identical with my pouring the poison' (1969, p.177). But since the pouring occurs before the death, we have the awkward result that the traveler is killed before he dies. Nonetheless, Davidson recommends that we should simply reconcile ourselves to this awkwardness—pretheoretic commonsense intuitions may be significant, but they aren't everything—and he suggests a variety of considerations that make it more palatable.[6] In the case of concept learning, a broadly analogous strategy might allow us to say that the learning does occur when the concept is first acquired (and before the ensuing conformation takes place). After all, concept learning is concept acquisition with a particular character; the fact that some elements that contribute to its character take place after the acquisition of the concept needn't be relevant. Granted, the concept learning case isn't entirely parallel with Davidson's, but the two are close enough, we think, to lend some credibility to Response 2. Just as the traveler may be said to be killed when the killer pours the poison (where the killing is partly determined by later events), so here a concept may be said to be learned when it is acquired (even though the learning is partly determined by later events).

Response 3. Suppose, however, that one were to insist that the learning of the concept doesn't take place until stage (3). Still, even then we needn't be moved by Fodor's charge that hypothesis testing models are circular. Though we might have to say that the concept is acquired before the learning is complete, that is only problematic on the assumption that the learning process must be completed at the same moment as the concept is acquired. But why think that? On the contrary, the far more natural way of viewing the situation is that the learning is itself a complex event that is stretched out in time and that it includes the acquisition of the concept as one of its components and the confirmation that it is the right concept as another of its components.[7] In that case, the concept may be acquired at t_1 while the learning may not be complete until t_2, but the learning includes t_1 and t_2 and much of what goes on between the two. Without the assumption that the moment at which concept learning is completed must be the moment at which the concept is acquired, there is no incompatibility between a concept being acquired at time t_1 and it being learned over the interval of time t_1- t_2. And so there is no circularity. Perhaps, then, Fodor is just slicing

[6] For example, he notes that some of the awkwardness might be due to associated epistemic issues rather than to the metaphysical claim itself—we can know about the pouring without knowing that it causes a death and so without knowing that the pouring is in fact a killing.

[7] Compare the example of murder, which presumably involves, among other things, both an intention to kill and a death. There is no reason to assume that the intention to kill and the death must be simultaneous in order for there to have been a murder.

up his events too thinly. The event of the complex concept being learned should be taken to incorporate both the processes that immediately account for the concept's acquisition and those that account for its justification. So learning a concept will typically take longer than, and hence not be simultaneous with, simply acquiring it.

It's time to take stock. One of the main considerations that underlies Fodor's new argument against concept learning is his skepticism about hypothesis testing models. We've been arguing that Fodor's charge of circularity is mistaken. First, a concept may be confirmed to be correct prior to its being explicitly represented, so that none of the justification necessary for learning need take place after the concept is acquired (as is illustrated by the justifiably confident dance student, section 3.1). However, even if we suppose that the concept to be learned is explicitly represented prior to confirming that it is the right concept, there are three defensible ways to avoid any threat of circularity. The concept might be taken to be learned when it is first explicitly formulated, owing to its acquisition having the right sort of environmental sensitivity, as in response 1. Alternatively, the concept might be taken to be learned when it is formulated, even though part of what makes this event a case of concept learning is the fact that it bears a particular relation to a later event of confirmation, as in response 2. Finally, the concept learning might be taken to be a complex event incorporating both the acquisition and the confirmation, which may be completed at a later time than the time at which the concept is acquired, as in response 3. All of these responses avoid the threat of circularity, and none are ruled out a priori.[8] Taken together, they demonstrate that the first premise of Fodor's new argument fails. Not only is it possible to learn complex concepts via hypothesis testing, but this type of concept learning is likely to be ubiquitous.

3.3. Beyond Hypothesis Testing

We now turn to the second part of Fodor's new argument—the claim that concept learning requires hypothesis testing. We'll argue in this section that Fodor is wrong about this point

---

[8] Recall that Fodor takes the argument in LOT2 to be an a priori argument against the possibility of concept learning (Fodor 2008, p. 130).

too. As with his first premise, however, seeing why we should reject it is illuminating about how concept learning actually works.

Fodor has always been emphatic that concept learning requires hypothesis testing. In LOT1, Fodor goes so far as to claim that 'there is only one kind of theory [hypothesis testing] that has ever been proposed for concept learning—indeed, there would seem to be only one kind of theory that is conceivable' (1975, p. 36). In LOT2, he says that hypothesis testing is 'the only candidate account' of concept learning (2008, p. 139) and that there is a consensus in cognitive science that concept learning is based on hypothesis testing. However, he also notes that 'though this consensus is pretty general, it's much more often than not inexplicit. There are very, very many theorists who accept HF without fully realizing that it's HF that they accept. I imagine, indeed, that that's the usual case' (2008, p. 132).

Whether Fodor is right about the cognitive science community's unvoiced commitments, we ought to ask why Fodor thinks that hypothesis testing models are all but inevitable. After explaining a hypothesis testing model in LOT2, Fodor considers the question, 'What's the evidence that children (for example) actually do learn concepts by some sort of induction?' (2008, p. 136). He responds to this question as follows (p. 136):

> Fair question… as far as anybody knows, there is simply no alternative. The only reliable way to infer from a batch of singular beliefs (*this instance of* EMERALD *is green; that instance of* EMERALD *is green; that other instance of* EMERALD *is green;* etc.) to general conclusions (EMERALD applies to *green things*) is to take the truth of the former as evidence for the truth of the latter. So either concept learning is what HF says it is or there isn't any such thing.

Unfortunately, this response isn't very helpful. The claim that there is no alternative 'as far as anyone knows' is controversial at best—as evidenced by the fact that, by Fodor's own admission, few in the cognitive science community see themselves as advocating hypothesis testing models of concept learning. And the remainder of the response *presupposes* that concept learning has the structure of an enumerative inductive inference. It may be that the only way to infer a general conclusion from a batch of singular beliefs is to take the singular beliefs as evidence for the general conclusion. But why should we suppose that concept learning takes this form to begin with?

Still, there is another, more general consideration that Fodor cites in LOT2—the one we noted in section 2—and this seems to be his driving motivation. He points to the need to distinguish genuine cases of learning from cases where a concept is acquired yet not learned, for instance, where a concept is acquired through 'surgical implantation', 'swallowing a pill', or 'hitting one's head' (2008, p 135). The problem is that whether a concept is learned or merely acquired isn't just a matter of whether the agent enters into a causal interaction with the environment. Causal interactions occur in both sorts of cases. So what distinguishes the instances where concepts are learned? For Fodor, it's that learning is a rational process in which the interactions with the environment provide the agent with evidence regarding the concept that she acquires. In contrast, surgical implantation and the like are, as Fodor likes to put it, brute-causal processes.

To this consideration, our initial reply is to note, with several of Fodor's other philosophical critics, that there is a considerable amount of logical space between brute-causal processes and explicit hypothesis testing (Samet & Flanagan 1989, Sterelny 1989). As these earlier discussions point out, neither the ordinary commonsense notion of learning nor its application in psychological research is restricted to cases of hypothesis testing. A cursory look at any introductory psychology textbook provides a wealth of examples where this is evident. Take, for instance, rote learning or the learning of facts. In these cases, information is recorded and cognitive processes such as rehearsal ensure that the information is retained for future use. But there is absolutely no reason to suppose that hypothesis projection and confirmation is required. In learning a phone number, for example, it's not as if one first has to hypothesize what the number is and then seek evidence to confirm that this hypothesis is correct.

We can see, then, that Fodor doesn't have much to back up the strong claim that hypothesis testing is mandatory for learning. But what isn't clear yet is how any of this translates into an alternative model of *concept* learning, as opposed to learning in other domains (e.g., fact learning or skill learning). In the remainder of this section, we will sketch a few ways in which complex concepts, in particular, can be learned that do not involve hypothesis testing. But first it is important to call attention to another crucial though implicit feature of Fodor's argument. In characterizing what is required for learning, Fodor moves swiftly from the need for a learner's activities to be rational (in some minimal sense) to the claim that her observations must count as evidence for what she acquires, and from here to the thought that the learner has to register the evidence by computing its bearing on an explicitly entertained hypothesis. While each step in this chain of inferences might be questioned, the point we want to call attention to at the moment is that Fodor is presupposing an *internalist* approach to justification. But an externalist approach would suffice just as well as a baseline requirement for explaining the 'rational' dimension of concept learning. In that case, what matters isn't so much that the agent recognizes the evidential value of what she observes, but rather that she employs cognitive mechanisms that deliver new concepts through processes that reliably reflect appropriate environmental contingencies. Whatever one thinks about epistemic justification more generally, there is a good case to be made that when it comes to 'the rationality' implicit in concept learning, externalist justification may well be all the justification that's required. In any case, it is useful to consider alternatives to Fodor's hypothesis testing that adopt an externalist criterion. With this preliminary point in mind, we will now present a sampling of different ways in which complex concepts can be learned without hypothesis testing.

Perceptual Learning. Consider what happens when someone forms a new complex concept as a result of perceiving an object or event that manifestly exhibits the combination of properties that the concept picks out. For instance, someone who encounters a black swan for the first time is likely to form the concept BLACK SWAN. She needn't have had the concept prior to her encounter, and it might never have occurred to her that non-white swans are a real possibility. But if she has the concepts BLACK and SWAN and sees a black swan with her own eyes, she will come into possession of BLACK SWAN and will be prepared to record and organize new information about these unexpected creatures. Learning a new concept in this way is largely a matter of perceiving one's surroundings and being open to the new arrangements they present. Moreover, the designation *learning* is perfectly apt. Acquiring BLACK SWAN upon seeing some black swans isn't anything like receiving a surgical implant. If we adopt the externalist approach to these matters, we can even say that the process is rational in the relevant, attenuated sense. After all, the various perceptual and cognitive processes that support the concept's acquisition reliably yield accurate descriptions of the objects and events with which they are causally interact. In fact, this is exactly what such systems are supposed to do. Perceptual-based learning of complex concepts is a paradigm of acquiring new concepts through the operation of psychological operations that have the function of extracting categorical information from the environment.

Communication-based learning. Concept learning can also be supported by explicit verbal instruction and communication. Think about what happens in a university classroom. In a logic course, for example, the instructor might convey the definition of validity by saying that a valid argument is one in which, if the premises are true, the conclusion must be true as well. A good attentive student might thereby *learn* the concept VALID ARGUMENT (or AN ARGUMENT IN WHICH, IF THE PREMISES OF THE ARGUMENT ARE TRUE, THE CONCLUSION MUST BE TRUE AS WELL). Of course, this learning will later be reinforced through examples and embedded into a broader knowledge of logical concepts, but the verbal communication alone can be sufficient for the student to have learned the concept from the teacher. What's more, learning via communication isn't by any means confined to the classroom or to concepts that are especially difficult to master. Think about the ordinary situation in which a friend describes how to make a new pasta sauce that you happen to be interested in. Hearing the verbal description all by itself allows you to construct a complex concept that represents all the steps in the recipe. And as with instances of learning by perceiving, these sorts of cases aren't anything like acquiring a concept through brain surgery. The language processing mechanisms that direct the concept's acquisition have the function of extracting information from the linguistic environment and of delivering new concepts corresponding to the linguistic input they are given.

Automatic associative learning. Finally, consider associative learning. Associative learning comes in many varieties and there are a number of different theories as to how it works. One type of associative process that presents an alternative to hypothesis testing is where a complex concept is automatically formed in response to the statistics of the mind's perceptual input. On this general approach, there are mechanisms that track perceptual properties and their statistical correlations and that construct novel complex concepts (according to the principles of a combinatorial semantics, not mere associative bonds) to encode these relationships. For instance, such a system might separately monitor the environment for pairs of properties, *F* and *G,* and be designed so that if it registers that a significant number of objects have both of these properties while few objects have only one of these properties on its own, then the system will form a new complex concept that encodes their conjunction (F&G). Once again, we have a cognitive system that has the function of extracting environmental information and that reliably yields environmentally relevant concepts. And, once again, the system's operations aren't at all brute-causal despite the lack of hypothesis testing.

What can Fodor say about these sorts of cases? Because he maintains that there are no alternatives to hypothesis testing for explaining how concepts are learned, he only has two options. (1) He can accept that these are examples of concept acquisition that don't involve hypothesis testing yet go on to deny that they count as genuine examples of learning. Alternatively, (2) he can insist that, despite appearances, our examples covertly involve hypothesis testing after all. Neither of these responses is especially plausible, however.

We have already touched on the first. The problem for Fodor is that the examples we have given are nothing like the brute-causal processes that he contrasts with genuine cases of learning. In most of the brute-causal scenarios that he mentions, it's little more than a lucky coincidence that the outcome is a new concept, much less one that is relevant to the agent's situation. (The coincidence isn't just lucky; it's so far-fetched that it's hard to take seriously, but we won't press that point here.) By contrast, consider once again what goes into learning a new concept by means of explicit instruction and verbal communication. The agent may be fortunate to have access to the right teacher or the right conversational partner, yet once this other person says what she has to say, luck drops out of it altogether. It's because the learner has the right cognitive equipment—including language processing systems—that she is able to arrive at the needed concept. This type of acquisition involves cognitive and perceptual processes that have the function of producing new complex representations on the basis of relevant information that is reliably extracted from the environment. It is nothing at all like acquiring a concept by miraculously being hit on the head in just the right way.

The more interesting response, and the one that Fodor shows some sympathy for, is to claim that our examples of learning hide a critical dimension in which hypothesis testing is going on. Recall Fodor's remark that 'there are very, very many theorists who accept

HF without fully realizing that it's HF that they accept' (2008, 132).[9] The problem with this response, however, is that there is no need for any hidden hypothesis testing, and no reason to believe that it occurs in the sorts of cases we have discussed. Consider again perception-based concept learning of the kind that is illustrated by the black swan example. In cases like this one, learning a new complex is a matter of assembling the concept that corresponds to a perceived object or event. Now in principle it could be that what happens is that the agent initially formulates a hypothesis concerning the identity of the concept in question—THAT THE CONCEPT THAT IS INSTANCED BEFORE ME IS THE CONCEPT *BLACK SWAN*—and then proceeds to test this hypothesis against further observations. But there is no need at all for this extra layer of highly reflective cogitation. What is far more plausible is that the visual system simply registers the presence of a black swan, automatically resulting in construction of the mental representation BLACK SWAN. This representation would then serve as a nexus for organizing further incoming information about the encountered animal or others like it.[10] Or consider the sort of case where someone learns a new concept through verbal communication. The learner hears her friend produce a verbal description and as a result manages to learn a concept of interest right there on the spot. Now our learner could, in principle, entertain various possible identity conditions for the concept and then seek to tease them apart through further observation. But it's unlikely that anything like this goes on in the normal case since the new concept would become available to the learner anyway as soon as the she understands the speaker's words. Nothing further needs to be done.

This concludes our critique of Fodor's argument as it applies to complex concepts. We have shown that hypothesis testing models are not, as Fodor has claimed, circular. There are cases of concept learning via hypothesis testing where there is no possibility of circularity because the full content of the hypothesis is confirmed prior to the explicit

[9] If Fodor were go this route, he'd surely add that our examples don't count as genuine cases of learning in the end, even with the covert hypothesis testing, since he maintains that hypothesis testing models are circular. But we have already dealt with the circularity charge.

[10] What if Fodor were to concede that the agent doesn't *explicitly* engage in hypothesis testing but were to insist that she does so *implicitly*. This won't help. If the process was merely implicit, then there would be no reason to suppose that BLACK SWAN was explicitly represented prior to its being learned, and Fodor's charge that hypothesis testing accounts are circular wouldn't get off the ground. Given his general strategy for arguing against concept learning, Fodor has to maintain that the hypothesis testing that goes into learning is explicit.

formulation of the complete hypothesis (as in the original dance example). In other cases, as in the modified dance example, there are several different ways to address the charge of circularity, any of which would suffice to show that complex concepts are learnable in such cases. We have also shown that hypothesis testing is not necessary for learning complex concepts anyway. In particular, we have outlined a variety of processes that do not involve hypothesis testing but that are well-suited to explaining how certain complex concepts are learned (perceptual learning, communication-based learning, and automatic associative learning). In the next section, we consider concept learning in relation to primitive concepts.

4. Learning Primitive Concepts

Though few theorists have endorsed Fodor's general anti-learning argument, many have agreed with his claim that *primitive* concepts can't be learned. This is because of the widespread commitment to what we call the *Building Blocks Model of Concept Learning*. According to this general approach, when a new concept is learned, it must be assembled from simpler concepts, and while these may in turn be assembled from yet simpler concepts (and so on), eventually all learned concepts are dependent upon an innate stock of primitive concepts from which they are composed. The innate primitive concepts amount to the fundamental building blocks of the conceptual system. In this section, we'll show that Fodor's skepticism about learning primitive concepts, and the Building Blocks Model that goes with it, are both misguided. Primitive concepts can be learned even though (being primitive) they can't be assembled from simpler concepts.

To see how learning a primitive concept is possible, the place to begin is with a theory of content for primitive concepts. This is because we need to be clear about the conditions that must be met for possessing these concepts. Given an explicit specification of these conditions, we can then ask how an agent's mind comes to satisfy them and whether there is any way that it could be done through learning. In earlier work, particularly in Margolis (1998), we developed a model of just this sort, based on Fodor's own theory of content—his *asymmetric dependence theory* (Fodor 1990). The model was designed to explain how an agent could learn natural kind concepts (e.g., concepts for different types of animals), treating these as primitive concepts. We chose Fodor's theory of content because it is specifically meant to cover primitive concepts, as Fodor devised the theory after coming to believe that most lexical concepts are primitive. Putting aside many of the details, the core idea of Fodor's theory of content is that a concept expresses the property that it is causally dependent upon in that instances of the property reliably cause the concept to be activated. For example, the concept ZEBRA expresses the property *zebra* because zebras reliably cause the activation of ZEBRA. That's not enough, of course, because things that aren't in the extension of a concept may cause it to be activated (e.g., if a horse is mistaken for a zebra). A second component of the theory deals with these cases by saying that the regularities that subsume them are dependent upon the causal link between zebras and ZEBRAS, and not the other way around. As Fodor would put it, the horse/ZEBRA regularity is asymmetrically dependent upon the zebra/ZEBRA regularity.

One important feature of Fodor's theory of content is that it analyzes content in terms of mind-world causal relations, abstracting away from the internal processes that occur within an agent's mind. So it allows for the possibility that two people can both possess ZEBRA and yet have very different beliefs and correspondingly different inferential dispositions regarding zebras. At the same time, it is important to remember that, even on such an account, the internal processes aren't irrelevant. This is because the mind-world relations that are constitutive of content have to be brought about somehow, and this will typically happen because of the various inferential connections that are associated with a concept. We call these systems of inferential connections *sustaining mechanisms*. Sustaining mechanisms establish and preserve the mind-world relations that constitute a concept's content.

Given the need for sustaining mechanisms, the question of how a primitive concept is acquired can be recast as the question of how one or more sustaining mechanism for the concept is acquired. While there are many different types of sustaining mechanisms, a useful point of orientation is to focus on one particularly interesting type, which we call a *syndrome-based sustaining mechanism*. This type of sustaining mechanism supports the possession of a natural kind concept by combining two fairly general sorts of cognitive dispositions. One tracks readily observable properties of a kind, while the other embodies the tendency to view the kind as having an essence or underlying reality that is shared by all of its members and that causally explains its observable properties. Together these dispositions allow for the construction of a sustaining mechanism for a concept like ZEBRA. In the standard case, encounters with zebras would cause an agent to record salient features of their appearance (their shape, motion, color markings, etc.) and associate these with a new representation (ZEBRA) that would become activated by things with much the same appearance. Appearances can be deceiving, however. Under some conditions, an ordinary horse might suffice to activate ZEBRA. But this is where the essentialist disposition does its work. The agent would be disposed to withdraw the judgment that something falls under the concept ZEBRA upon learning information that would indicate the absence of the right essence, making the horse/ ZEBRA connection less basic than (and thus asymmetrically dependent on) the zebra/ZEBRA connection.

There are two key points to notice regarding this proposed model. First, it doesn't implicate any Fodorian hypothesis testing. To acquire ZEBRA, the agent needn't confirm a hypothesis about the individuating conditions on the concept ZEBRA. Indeed, she may never come to explicitly represent individuating conditions for the concept at all. Second, despite the lack of hypothesis testing, there is every reason to believe that the model is still a learning model. For one thing, it doesn't imply that ZEBRA and the like are simply innate. It's not as if the model says that a sustaining mechanism for ZEBRA is all wired up in advance and simply waiting for an innately specified triggering condition to cause it to become activated. Far from it. What is innate, according to the model, is a general cognitive organization for creating a range of syndrome-based sustaining mechanisms in response to new natural kinds. The reason this organization leads to the creation of a sustaining mechanism for ZEBRA, and thereby explains the acquisition of ZEBRA, is because of the cognitive processing that this organization initiates given causal contact with zebras. If the same organization were brought to bear in encounters with other animals (e.g., lions or giraffes), then the result would be concepts for these other animals instead. As with the learning models for complex concepts discussed earlier, the mechanisms involved in the acquisition of the concept have the function of producing new representations on the basis of relevant information that is systematically and reliably extracted from the experienced environment. And again, like the learning models for complex concepts discussed earlier, the construction of a syndrome-based sustaining mechanism, as we are envisioning it, stands in clear contrast with Fodor's brute-causal processes (futuristic neurosurgery, a miraculous blow to the head of just the right sort, etc.).

It would be nice if we could end things here, and in a past discussion, we were content to argue that a model along these lines suffices to show that even primitive concepts can be learned (Laurence & Margolis 2002). But Fodor will have none of this. In LOT2, he offers a number of objections that are meant to show that our model fails to illustrate how learning a primitive concept is possible. Fortunately, the objections in LOT2 aren't successful, and it only strengthens our own case against the Building Blocks Model to see why they don't work.

The first objection challenges our use of Fodor's asymmetric-dependence theory and its commitment to an externalist treatment of conceptual content. Fodor charges that we 'overstate the case for semantic referentialism' (2008, p. 141), as his argument against concept learning is neutral about the type of semantic content that concepts have. The entirely general form of argument that Fodor claims to be using comes out just a bit earlier in the text. In a discussion couched in the form of a dialogue, he asks what makes it the case that an agent learns one of a pair of co-extensive concepts rather than the other.
Fodor's answer (2008, pp. 134-5):

> I mean, mustn't it be something like this: in consequence of their respective experiences, one learner comes to think to himself 'All those things are Cs' but the other comes to think to himself 'all those things are C*s'? … And is not inductive inference the process par excellence by which one proceeds from representing some things as Cs to representing all such things as Cs? … And is not the formation and confirmation of hypotheses the very essence of inductive inference? … So, does not the fact that it is possible to learn one but not the other of two distinct but coextensive concepts show that concept learning is indeed some kind of hypothesis testing?

But as we remarked in section 3.3, it shouldn't be taken for granted that concept learning amounts to inferring a general belief from various singular beliefs and hence that learning must take the form of an enumerative inductive inference. Fodor isn't arguing that concept learning requires hypothesis testing. He is simply presupposing that this is all it can be.

Still, the attention given to co-extensive concepts suggests a more pointed criticism of our model. This is that our sustaining mechanism account of concept acquisition is problematic since it can't distinguish C from C*. In response to *this* worry, we should first mention that in earlier work we were at pains to emphasize that our use of Fodor's asymmetric-dependence theory is only for illustrative purposes and that we are not committed to this particular theory of content (Laurence & Margolis, 2002).[11] But regardless, the challenge pertaining to the learning of coextensive concepts is one that advocates of the asymmetric-dependence theory (and semantic externalists more generally) can readily handle without having to capitulate to Fodor on the question of whether learning requires hypothesis testing. All that an externalist needs to do to address the problem of coextensive concepts is to maintain that concept identity isn't solely a matter of extension. When Fodor isn't arguing about whether concepts can be learned, he himself has been clear that externalists can avail themselves of the notion of a mode of presentation. For Fodor, modes of presentations are realized by the formal properties of the mental representations in which thinking takes place (Fodor 2008, ch. 3). Alternatively, externalists can say that a concept's identity is partly constituted by its conceptual role whether or not conceptual role is taken to be part of a concept's content (Margolis & Laurence, 2007). Thus it's open to an externalist to say that what makes it the case that one acquires C as opposed to the coextensive C* is that the process of acquisition in the first case results in a representation that is partly constituted by the C-formal or C-conceptual-role properties, together with C's extension. Assuming that the acquisition involves a process akin to the construction of a syndrome-based sustaining mechanism, Fodorian hypothesis testing needn't come into it.

Fodor's second objection goes right to the heart of our earlier model by challenging how it transforms the question about concept learning into a question about the way that sustaining mechanisms are acquired. As Fodor puts it, 'If the sort of referentialist/atomist story about conceptual content that Margolis and I like is true … then learning a theory can be (causally) sufficient for acquiring a concept. But it doesn't follow that you can learn a concept. So here we are, back where we started; we still don't have a clue what it might be to learn a concept' (2008, p. 144). This objection requires a bit of unpacking. In LOT2,

---

[11] Though we do not have space to argue for the claim here, we believe that analogous models can be developed for other theories of content and also for a variety of other types of primitive concepts besides natural kind concepts (see Laurence & Margolis, in prep.).

Fodor cites a pair of related considerations that are supposed to show that acquiring a sustaining mechanism and learning a concept are entirely different things. The first is that even if a sustaining mechanism is learned, there is no guarantee that the concept it gives rise to is learned as a result (2008, p. 144):

> … 'You can learn (not just acquire) A' and 'Learning A is sufficient for acquiring B' just doesn't imply 'You can learn B'. For, the following would seem to be a live option: If you acquire a concept by learning a theory, then something is learned (namely, the theory) and something is (merely) acquired (namely, the concept); but what is learned isn't (merely) acquired and what is (merely) acquired isn't learned. To acquire the concept C is to lock to the property that Cs have in common; and such lockings may be mediated by theories. The theory that mediates the locking between the concept and the property that the concept is locked to may be, but needn't be, rational, or coherent, or well evidenced, to say nothing of true. That's why Ancient Greeks, who thought stars were holes in the fabric of the heavens, could nevertheless think about stars.

Fodor uses the term 'theory' as shorthand for what we are calling a sustaining mechanism. In the sort of case that he is imagining, though the sustaining mechanism is learned, it turns out to be a causal fluke that the sustaining mechanism provides semantic access to the property that the concept expresses. Much of the information in the sustaining mechanism is false and the sustaining mechanism is irrational or lacking in coherence and evidential support. But if it is, as it were, an accident that the sustaining mechanism mediates access to the property, how can the agent be said to have learned the concept?

The second, related consideration Fodor raises is meant to increase the gap between acquiring a sustaining mechanism and learning a concept. It does this by reintroducing a familiar worry (2008, p. 144):

> [Y]ou can [also] acquire a concept by *acquiring* a theory (i.e., by acquiring it but not learning it). I'm dropped on my head and thereby acquire the geocentric theory of planetary motion, and thereby become linked to, say, the property of being a planet. In such cases, neither the theory I've acquired nor the concept I've acquired has been learned.

This time the problem isn't that it's odd and unexpected that the sustaining mechanism provides semantic access to the property its concept expresses. It's that the way that the sustaining mechanism itself is introduced is the result of a causal fluke. But if the sustaining mechanism isn't learned, if it just pops into someone's head, how can the concept that it supports be learned as a consequence? Taken together these two worries are intended to show that acquiring a sustaining mechanism can't be sufficient for learning a concept and consequently that the attention we have given to sustaining mechanisms is wrongheaded.

Neither of these considerations succeeds at undermining our account as a learning model. Indeed, whatever plausibility they have as counterexamples to our model of concept learning rests on a misunderstanding of our model and the dialectic it is meant to address. Recall that Fodor's argument against concept learning is intended to establish that it is *impossible* to learn any new concept. So in response to our model of concept learning, it is no good for Fodor to argue that some cases of acquiring a sustaining mechanism fail to constitute learning a concept. He needs to show that *every* case of acquiring a sustaining mechanism fails to constitute learning a concept—that there aren't *any cases at all* where acquiring a sustaining mechanism would count as learning a concept. One cannot possibly show that it's impossible to learn a concept through acquiring a sustaining mechanism by showing that a few highly idiosyncratic instances of this sort fail to count as learning. One might as well argue that since penguins can't fly, it's impossible for birds to fly.

Our argument against Fodor's case for the impossibility of concept learning centered on a specific instantiation of our general sustaining mechanism account. (We never claimed that acquiring a sustaining mechanism for a concept constitutes concept learning no matter what the sustaining mechanism is like or how it is acquired). Accordingly, for Fodor to rebut our argument, it is incumbent upon him to show that a concept could not be learned *in the sort of case we outlined*—a case much like the one we described above in introducing the idea of a sustaining mechanism. Here the learner acquires the sustaining mechanism for a new concept, for example ZEBRA, directly through ordinary perceptual causal contact with instances of the concept, accurately recording relevant observable properties of the kind, under the direction of a general intention to learn about this new kind. It is not a fluke that the information in the sustaining mechanism establishes a nomic dependence between the concept and its instances, nor is it merely a fortuitious quirk that the sustaining mechanism is acquired at all. The information recorded is relevant information that is acquired through perceptual and cognitive processes that have the function of recording such information and of organizing it into new representations of natural kinds. To address our challenge, Fodor must demonstrate that it is impossible to learn a new concept when a syndrome-based sustaining mechanism is acquired in this sort of paradigmatic case; Fodor's flukey cases are simply irrelevant given a proper understanding of the dialectic.

Fodor's third and final objection raises the question of whether we are warranted in describing the syndrome-based sustaining mechanism as one in which concepts are truly *learned*. Throughout his discussion in LOT2, Fodor maintains that hypothesis testing is the only conceivable model of concept learning. Accordingly, Fodor should be expected to resist our claim that concepts can be genuinely *learned*, as opposed to merely acquired, via a process of this kind. In fact, in a recent conference that was dedicated to Fodor's views on concept acquisition, Fodor not only sounded as if he took it to be an a priori truth that learning requires hypothesis testing but that our model should be rejected barring an alternative *definition* of learning (i.e., a definition that addresses the motivations that originally prompted the hypothesis testing analysis and that can serve as a principled guide for identifying when learning occurs).[12] We confess that we don't have a definition to offer. But it would be deeply ironic if Fodor, of all people, held that against us. A central theme in Fodor's work is that the search for definitions is almost always futile. Fodor has been especially critical of the idea that lexical concepts, in particular, can be defined.[13] But if most ordinary lexical concepts can't be defined, it's hardly fair to ask us to provide a definition of LEARNING. Instead, the assumption ought to be that that LEARNING *cannot* be defined.

What's more, a brief inventory of different types of learning suggests that LEARNING picks out a rather heterogeneous set of phenomena. Consider the diversity that is manifestly associated with rote learning, learning a language, learning a complex manual procedure (e.g., how to play the violin), learning the contents of a room, learning a novel route to an old location, learning algebra at school, learning what an avocado tastes like, learning which kinds of animals are dangerous, and learning an implicit cultural norm.

---

[12] *Symposium on Solutions to Fodor's Puzzle of Concept Acquisition*, Annual Cognitive Science Society (2005), meeting held in Stressa, Italy. The transcript is available online: <http://www.wjh.harvard.edu/~lds/pdfs/Niyogi_Snedeker-2005.pdf>

[13] In 'The Present Status of the Innateness Controversy', Fodor notes, 'I once heard Professor Gilbert Harman remark that it would be *surprising* if "know" were definable, since nothing else is. Precisely' (Fodor 1981, p. 285). See also Fodor *et al*. (1980), which is aptly titled 'Against Definitions'.

These diverse phenomena are all natural to describe as cases of learning, but there is no reason to suppose that the underlying processes share a common defining set of features. (They certainly don't all involve hypothesis testing, which they should if Fodor views about learning were correct.) To the extent that the scientific community is beginning to understand what goes on with these and other cases of learning, it is becoming increasingly plausible that specialized learning mechanisms are often responsible in good part for our abilities in different task domains. For example, much of language acquisition depends on domain-specific learning mechanisms, so there is no reason to suppose that learning English is accomplished in quite the same way as, say, learning a new route home. Even within a given task domain, the same outcome can depend upon rather different mechanisms. One could learn the contents of a room by opening the door and looking inside, but also by taking advantage of someone else's testimony. Though the result may be much the same, the mechanisms involved are strikingly different. Or to take another example, one could learn a route home by exploiting memorized landmarks but also by relying on dead reckoning—again, very different underlying mechanisms. The more we discover about the mind, the more we have to face the fact that there may be very little that all cases of learning invariably have in common (Gallistel 2000).

Still, perhaps something can be said about some of the characteristics of typical instances of learning. This isn't to give a definition of learning, just to note a few of the features that implicitly guide the recognition of certain clear-cut cases. We'd suggest three.[14] The first and most basic is that learning generally involves a cognitive change as a response to causal interactions with the environment. Of course, not all changes that trace back to an organism's environment will count as learning. That is the point of Fodor's examples where concepts are acquired through futuristic neurosurgery or through a miraculously fortuitous hit on the head. And not all cases of learning will involve environmental sensitivity, as some learning may be wholly a priori. Nevertheless, one important feature of learning is that it often, perhaps typically, involves a sensitivity to the environment. The other two features we wish to suggest are ones that take things a little bit further by highlighting some aspects of the causal interaction that occur in paradigmatic cases of learning. One is that learning often implicates a cognitive system that isn't just

[14] For a related discussion of the general characteristics of learning, see Weiskopf (2008).

altered by the environment but, in some sense, has the function to respond as it does. For example, learning facts about the locations of various objects when entering a room isn't just a matter of having your mind altered upon perceiving the situation. The changes presumably are of the sort that our perceptual systems and related belief-fixation mechanisms are designed to subserve. In contrast, when you get hit in the head, as in Fodor's example, your mind might be miraculously altered in a useful way, but the intervening mechanisms don't have the function of subserving these sorts of changes; it's just a matter of blind luck. The other suggestion is that learning processes are ones that connect the content of an experience with the content of what is learned. The two aren't merely causally related. They are *semantically* related. Hypothesis testing exemplifies one type of semantic relatedness, but it hardly exhausts the possibilities. For example, the rote learning of a list of numbers may involve reciting the numbers several times out loud, chunking the numbers in thought, and associating the numbers with other memorable items (famous dates, a friend's birthday, etc.). By any reasonable standard, the processes that are integral to these activities are ones in which the outcome is semantically related to the preceding experience. It's not as if the list enters into memory through sheer coincidence. The cognitive processes that bring about the change in thought are ones that undoubtedly turn on the contents of the mental states involved in the transition.

With this brief characterization of learning in hand, we can return to the claim that paradigmatic instances of acquisition involving the syndrome-based sustaining mechanism model are worthy of being described as learning. The model clearly has all three of the features that we have identified. In these sorts of paradigmatic cases, the learner gathers information about the kind based on her perception of the members she encounters, so there is no question that the change is grounded in causal interactions with the environment. Also, the information that is presented in experience isn't capable of directly creating a syndrome-based sustaining mechanism all by itself. The gathering of this information is guided by expectations about natural kinds in general and by expectations that derive from previous experiences with members of the kind in question and with members of similar kinds. It is filtered and processed by cognitive operations that take perceptual information and use it to control the application of a new concept, including its application in the context of similar future experiences. Here we have about as clear a case of semantically relevant processes as one could want. Finally, it's perfectly reasonable to suppose that the cognitive mechanisms and dispositions that support the formation of syndrome-based sustaining mechanisms have the function of building them as they do. It's even plausible that the essentialist disposition is owing to an adaptation for interacting with natural kinds. In short, our model of concept learning manifestly exemplifies the pre-theoretic understanding of learning, satisfying the characteristics of typical instances of learning extremely well.

We conclude that the case that primitive concepts can be learned is quite strong. None of Fodor's responses to our model are successful. First, though our model isn't wedded to an externalist theory of content, it wouldn't matter if it was, since co-extensive concepts can be teased apart by appealing to a concept's mode of presentation. Second, the existence of exotic cases involving causal flukes could not possibly show that it is impossible to learn concepts via our model. Such cases do nothing to address the paradigmatic cases that should be at issue in evaluating our model; they are simply irrelevant. Third, there is no need for us to provide a definition of learning in order to claim that our model is a learning model. It is enough if the model satisfies the features of paradigmatic cases of learning, which it does.

Earlier we established that complex concepts can be learned. We can now add that primitive concepts can be learned too. Taken together these conclusions thoroughly undermine Fodor's skepticism about learning new concepts. The conclusion regarding primitive concepts is particularly interesting, however, because many theorists who do not share Fodor's views regarding the impossibility of concept learning nonetheless hold that primitive concepts cannot be learned. These theorists, who are proponents of the Building Blocks Model, are committed to the view that there are significant limitations to the range of thoughts that human beings can entertain. Steven Pinker offers a useful analogy for envisioning the situation. We can think of the mind as a typewriter. The many strings of symbols that a typewriter can produce aren't specifically encoded in the machine. They are a product of a small number of characters that are part of the typewriter's fixed architecture. But all the same, these characters and their various combinations establish a limit that the typewriter can never overcome. 'Type … all you want; though you can bang out any number of English words and sentences and paragraphs, you'll never see a single character of Hebrew or Tamil or Japanese' (Pinker 2007, p. 93). The point, in other words, is that the expressive power of the conceptual system is fixed given its principles of combination and its innate primitives. The concepts that an agent possesses at any given time has to do with the combinations that have actually taken place, but no matter how much thinking she does, she is always stuck within the space of concepts that is imposed by these two parameters. What our argument that primitive concepts can be learned shows, however, is that this tempting picture of the mind is substantially mistaken. New primitive concepts can be learned and so it *is* possible to fundamentally increase the expressive power of the conceptual system. Human beings can entertain thoughts that go beyond those constructible via a compositional semantics from the innate stock of primitive concepts.

5. Fodor's Biological Account of Concept Acquisition

We have not yet considered Fodor's positive account of concept acquisition. In addition to the new argument against concept learning in LOT2, Fodor has also substantially updated his account of concept acquisition. This is of interest not only because proponents of concept learning can shift the burden of proof and ask him how he proposes to explain where our concepts come from, but also because Fodor opts for the unusual strategy of claiming that concept acquisition is essentially a noncognitive *biological* phenomenon. Seeing how far a biological account can go reveals some important lessons about concepts and how we come to possess them.

To appreciate the novelty of Fodor's positive account of concept acquisition, it helps to back up and see how it contrasts with Fodor's earlier thinking about these matters. In LOT1 (and in Fodor 1981), Fodor had taken his argument against learning to show that most lexical concepts are innate. In subsequent work, however, Fodor has become hesitant to conclude that any concepts are innate and has proposed that we need to circumvent the learned/innate dichotomy. The way to do this, he thinks, is to tell the story about concept acquisition not at the cognitive level—the level of intentional states and processes—but at the neurological level (Fodor 1998, p. 143):

> [T]hough there has to be a story to tell about the structural requirements for acquiring DOORKNOB, intentional vocabulary isn't required to tell it. In which case, it isn't part

> of cognitive psychology. Not even of "cognitive neuropsychology" … (as opposed, as it were, to neuropsychology *tout court*).

On this view, concept acquisition is a wholly noncognitive affair that is to be explained directly and entirely in neurological terms. Notice that the claim here isn't simply the uncontroversial idea that psychological states are realized in the brain and ultimately dependent upon the brain's activities. Rather, Fodor is making the far stronger claim that concept acquisition is subject *only* to neurological explanation. If Fodor is right, cognitive psychology is no more relevant to the study of concept acquisition than it is to the study of red blood cell production or digestion.[15]

At first, this may not seem like a tenable approach to concept acquisition for a reason that Fodor himself was quick to identify. Nonpsychological accounts of concept acquisition seem to make a mystery of the fact that concepts are regularly acquired through exposure to their instances. There seems to be no reason why a nonrational, noncognitive neurological process should require exposure to doorknobs to acquire the concept DOORKNOB, or why such a process would lead to the production of the concept DOORKNOB on exposure to doorknobs (as opposed to GIRAFFE or some other unrelated concept). The solution can't be that doorknobs offer opportunities for confirming hypotheses about DOORKNOB or for representing their salient features. That would be to resort to psychology and its intentional states.[16] Fodor (1998) refers to this problem as the *doorknob/DOORKNOB problem* and offers to solve the problem with a bold metaphysical theory. The reason someone acquires DOORKNOB when interacting with doorknobs is not because of their psychology, but rather because of the metaphysical nature of the property of being a *doorknob* (i.e., the referent of DOORKNOB). According to Fodor's metaphysical theory, the property of being a doorknob is partly constituted by the fact that it leads to the acquisition

[15] Though our focus throughout this paper is on concept learning, it is worth noting that Fodor's anti-cognitivist account is not only opposed to concept learning models but to *all* cognitive-level accounts of concept acquisition, even those which do not claim to involve learning (e.g., acquiring a concept through combining concepts one already has in imagination).

[16] Regardless of whether learning involves hypothesis testing, it ought to be clear that learning models in general don't face this problem. This is because learning models will take doorknobs to be a source of information about doorknobs and hold that learning mechanisms are positioned to extract that information. By contrast, doorknobs don't provide any information about giraffes, so it's hardly surprising that you don't get GIRAFFE when you see a doorknob. The doorknob/DOORKNOB problem only arises because Fodor has rejected all cognitive-level stories about concept acquisition.

of DOORKNOB. Since it is in the nature of what it is to be a doorknob that, under certain conditions, we react to doorknobs by thinking DOORKNOB, then, Fodor claimed, it should no longer be mysterious that doorknobs are implicated in the acquisition of DOORKNOB.

Now let's fast forward to LOT2. Fodor's LOT2 theory of concept acquisition retains the core elements in Fodor (1998). There is still the insistence that we should deny both learning and innateness and hold out for a nonpsychological theory. LOT2 also continues to endorse the metaphysical solution to the doorknob/DOORKNOB problem. What is distinctive about the account in LOT2 is that Fodor now takes concept acquisition to proceed in two stages. During the first stage, a stereotype for a concept is learned.[17] Importantly, the stereotype isn't identical with the concept. However, the fact that stereotypes are learned in acquiring concepts partly explains why concepts appear to be learned when, according to Fodor, they aren't. The appearance is owing to the fact that a stereotype is related to its concept, and while the concept itself isn't learned, the stereotype is. The second stage of concept acquisition on Fodor's new model occurs once the stereotype is in place. As Fodor sees it, a neurological process takes over and generates the concept from its stereotype. Crucially, the second stage is noncognitive. It is '*sub*intentional and *sub*computational… a kind of thing that our sort of brain just does' (2008, p. 152). The second stage in Fodor's account is supposed to be where most of the action is and is why Fodor maintains that concepts aren't learned. It's because of our biological makeup that we arrive at the right concept corresponding to a stereotype; learning has nothing to do with it.

Fodor summarizes his LOT2 account of concept acquisition as follows: 'here's my story about concept acquisition: What's learned (not just acquired) are stereotypes (statistical representations of experience)' (2008, p. 162). Once you learn a stereotype, then noncognitive, nonrational neurological processes get you to the concept: 'in particular, it's a brute fact about the kind of animals we are (presumably about the kind of brains we have); and it's the bedrock on which the phenomenon of concept acquisition rests' (2008, p. 161).

---

[17] For present purposes, we can take a stereotype to be a complex representation that is associated with a concept. A concept's stereotype encodes information about properties that instances of the concept tend to possess.

This account of concept acquisition faces a number of serious problems. Let's begin by looking at the first stage in the acquisition process—the stage where stereotypes (but not their concepts) are learned. The main problem here is that the claim that stereotypes are learned is incompatible with Fodor's argument against concept learning. Our objection can be put as a dilemma. The first horn springs from Fodor's insistence that learning requires hypothesis testing. Assuming that it does, learning a concept's stereotype would necessitate putting forward hypotheses about the stereotype's individuating conditions and testing these against relevant observations. But then, following Fodor's own logic, putting forward the correct hypothesis would require having the stereotype prior to learning it, which would entail that the stereotype can't really be learned after all. Now it's true that we have rejected Fodor's circularity argument, but the present criticism concerns the internal coherence of Fodor's position. Perhaps Fodor could avoid the charge of circularity if he were to allow that not all learning reduces to hypothesis testing and were to claim that stereotypes are learned in some other unspecified way. But then he would face the second horn of the dilemma: this qualification would undermine the circularity argument against *concept* learning. Were Fodor to agree that learning doesn't necessitate hypothesis testing, there is nothing to stop his critics from countering by proposing that concepts are learned in this other unspecified way too. In short, if Fodor's argument against concept learning is sound, then it undermines his own account of concept acquisition, and, by the same token, if his biological account of concept acquisition is viable, it undermines his argument against concept learning.

Perhaps the most fundamental problems for Fodor's new account, however, stem from a family of considerations that provide grounds for thinking that concept acquisition requires a psychological treatment. Fodor himself anticipates this challenge somewhat in the form of his doorknob/DOORKNOB problem, but he fails to appreciate the scope of the problem. For example, consider the fact that different people will often acquire different concepts on exposure to the same physical environment because they have varying interests in the same segment of the world. Dog enthusiasts have concepts corresponding to dozens of breeds, while people who are indifferent to dogs often have just a handful. Similarly, surfers have concepts corresponding to numerous types of waves and surfing conditions[18],

---

[18] See, e.g., the surfing lingo website <http://riptionary.com>.

while most non-surfers have only a few generic concepts like CHOPPY WATER and SMALL WAVES. What's more, these examples aren't solely about the import of varying interests. They illustrate the significance of the surrounding culture. Dog enthusiasts and surfers hang out with others like them and pick up on the cultural norms of the groups they identify with. These interactions can be just as important as the interactions with the aspects of the physical world that the group cares about. But how can a purely biological process account for this fact? If mutable socially propagated norms are what matters, we need a mechanism that is calibrated to the *social* world. It's hard to see how this could be anything but a psychological mechanism, one that is chockfull of intentional states and processes.

As it happens, many concepts reflect the surrounding culture and it matters a great deal which culture it is that a learner grows up in. For example, Medieval Europeans conceptualized health and disease in terms of humors—bodily substances that, according to the theories of the time, need to be kept in balance with one another. Few contemporary Europeans have these concepts, though they have as much exposure to instances of good and bad health as their historical counterparts. Or take the Newtonian concept of GRAVITY. People who haven't been exposed to post-Newtonian physics aren't in a position to acquire this concept even though they experience the same sorts of causal interactions that exemplify gravitational influence (falling rocks, tides, etc.). Likewise for the logical concept of VALIDITY, which is evidenced by all the valid arguments one is exposed to, but typically is acquired only after a university course in logic. Cultural forces are even more significant in cases where there is *no* physical manifestation of the items the concepts are supposed to refer to. The Roman Catholic concept of PURGATORY, for example, is of a place, or state of being, that no living person has actually experienced. That concept comes from cultural products—books, stories, sermons, etc.—that can only have their influence through psychological processes that extract their meaning.

One way that Fodor might try to mitigate the impact of cases like these is by claiming that the reason why people in different cultures end up with different concepts is that they learn different stereotypes. This response emphasizes the fact that the second stage of concept acquisition—the crucial biological stage—can't occur until the right stereotype in place. While this response might help with some of the cases, it faces three serious problems.

First, even when agents have access to the *same* stereotype, the surrounding culture can have a profound effect on how the world comes to be conceptualized. Color concepts are a well-known case in point. Though people all around the world are equipped with essentially the same sensory systems, there is a significant amount of variation in the basic color concepts found in different cultures (Davidoff et al. 1999). The variation doesn't trace back to differences in surface reflectances or to other physical properties that are present in the environment. Rather, it's largely a matter of how different cultures have come to establish and encode the boundaries of their categories. Children inherit the local way of drawing these boundaries during the course of learning their language. What can Fodor say about this? His neurological account of concept acquisition stipulates that once the stereotype is in place, biology takes over and delivers the concept. But with color concepts, there aren't different stereotypes in different cultures. Focal instances of basic color concepts are highly similar across cultures; it's the breadth and boundaries of the concepts that differ. So there's nothing in Fodor's account to explain why children in different cultures end up with concepts that match their own community's way of doing things. If anything, Fodor's account predicts that children across the globe should end up with exactly the same color concepts since the same biological principles would be activated by much the same stereotypes—a prediction that unfortunately doesn't stand up to the facts. Much the same point can be made in light of the fact that our stereotypes for concepts are often highly impoverished and thus are unlikely to differ for related but distinct concepts. For example, for many people the stereotype for GERBIL is essentially the same as for HAMSTER. But this needn't stop them from acquiring one of these concepts but not the other (or even from acquiring both) on the basis of the very same stereotype. Fodor's view does nothing to relieve the mystery surrounding such differential concept acquisition.

The second problem with trying to use stereotype learning to explain away psychologically-mediated concept acquisition is perhaps even more serious. Many concepts can be acquired in the absence of any stereotype at all. In fact, Fodor himself has argued that complex concepts typically don't have stereotypes (Fodor 1981, 1998). But clearly, if a concept doesn't have a stereotype, then variable concept acquisition can't be explained by stereotypes. In particular, for all concepts that lack stereotypes, we are left with just the noncognitive biological part of Fodor's story about concept acquisition. And this part of Fodor's story has nothing to say to the evidence suggesting that concept acquisition is psychologically mediated. Consider, for example, the concept AN INVESTMENT FUND OPEN TO A LIMITED RANGE OF INVESTORS THAT UNDERTAKES A WIDER RANGE OF INVESTMENT AND TRADING ACTIVITIES THAN LONG-ONLY INVESTMENT FUNDS, AND THAT, IN GENERAL, PAYS A PERFORMANCE FEE TO ITS INVESTMENT MANAGER.[19] Prior to learning this concept (e.g., through verbal instruction), one is highly unlikely to have a stereotype associated with it. The same will be true of endlessly many of complex concepts (e.g., the Scottish Country Dance concept from section 3). It is also likely to be true of lexical concepts with highly theoretical content, at least for many nonexperts—MOLECULE, ARGON, ACETYLCHOLINE. These are not concepts that we learn by first learning a stereotype for them.[20]

This brings us to the third problem with the attempt to explain away psychologically mediated concept acquisition by appeal to stereotype learning. Many concepts are learned via the operation of psychological processes that go beyond stereotype formation. For example, some are learned alongside theories that they are embedded in (e.g., GRAVITY), and whether they are learned turns not on whether a stereotype for them is learned but on whether the learner is exposed to the relevant theory. In some cases, such as with natural kind concepts, there is arguably a default system devoted to gathering particular types of information about the kind and processing it in accordance with domain-specific inferential patterns. We can even predict which kinds of concepts an agent is likely to form based on considerations about the representational processes underlying concept acquisition in that domain. This makes sense if the concepts are acquired on the basis of the representational processes that support these predictions, but it is nothing short of a mystery on Fodor's biological account. Consider, for example, Pascal Boyer's analysis of the origins of concepts of spiritual beings. Boyer notes that the full range of possible supernatural concepts is far larger and more varied than what is actually found across cultures and consequently that supernatural concepts can't simply be a matter of generating new

---

[19] <http://en.wikipedia.org/wiki/Hedge_fund>.

[20] There is room for debate about whether concepts like MOLECULE or the concept corresponding to the dance have stereotypes (Prinz 2004, Fodor 1981). For the present point, all the really matters is that these stereotypes, if they exist at all, are so anemic that they cannot do the work that Fodor requires of them.

representations for strange incidents. Boyer's theory predicts, instead, that the most intuitive supernatural concepts are rooted in innate systems of inference (e.g., folk biology) and depend upon isolated counterintuitive deviations from the normal case (e.g., trees that talk), which make them memorable and apt for cultural transmission. This predication fits well with the anthropological record and with experiments that have been designed to test just these sorts of effects on memory (Boyer 2001).

In sum, Fodor's positive theory of concept acquisition faces a number of serious challenges. First, he faces a dilemma stemming from stage 1 of the acquisition process, which involves stereotype learning. Either his argument against concept learning undermines this aspect of his positive account or else his account of stereotype learning undermines his argument against concept learning. His account also fails to adequately explain away the robust and diverse evidence for maintaining that concept acquisition is a psychological-level phenomenon. This evidence includes the doorknob/DOORKNOB problem, but the doorknob/DOORKNOB problem is just the tip of the iceberg. More important is the fact that concept acquisition depends profoundly on one's *cultural* environment, as mediated by one's psychology. Both the number of concepts one acquires about a given domain (e.g., dogs, waves) and which specific concepts one acquires (e.g., color), depend on one's cultural environment. And the dependence is often very deep, as illustrated by the many concepts which are largely cultural, with little or no grounding in one's immediate environment (e.g., PURGATORY). Moreover, the variation cannot be explained by stereotype differences given that the conceptual variation is possible without variation in stereotypes—many concepts either lack stereotypes altogether or else fail to have sufficiently robust stereotypes to discriminate between related concepts.

In stark contrast to Fodor's biological view, accounts of concept acquisition that appeal to learning do not face any of these problems. The stereotype-learning dilemma disappears since concept learning accounts don't deny that either stereotypes or concepts can be learned. The doorknob/DOORKNOB problem has a straightforward and satisfying solution: concepts are often acquired through encounters with their instances because they are learned at least partly on the basis of collecting information about their instances. And the problems stemming from cultural embeddedness never arise, since concept learning accounts happily accept that many aspects of one's cultural surround are represented and feed into concept learning.

In retrospect, it should hardly be surprising that Fodor's new view of concept acquisition would face so many difficulties. Fodor's biological account should have been suspect from the start since its anti-cognitivism flies in the face of the deepest motivations behind the cognitive sciences—motivations that go back to origins of cognitive science in its opposition to eliminativist-behaviorism. This becomes apparent if we consider the analog of Fodor's theory of concept acquisition in the realm of language. A theory of language acquisition along the lines of Fodor's biological account of concept acquisition would hold that language is neither learned nor innate. The core processes involved in language acquisition, according to this sort of account, are nonrational, noncognitive, neurological processes; language acquisition is simply 'a brute fact about the kind of animals we are (presumably about the kind of brains we have)', and does not admit of a cognitive level explanation. Moreover, the reason why people who are exposed to English acquire the ability to speak English (as opposed to Italian, Mandarin, or the ability to play the violin, for that matter) is because English sentences are instances mind-dependent types — to be an English sentence just is to be the kind of thing that makes minds like ours jump to having the capacity to understand English. If only the science of the mind were so easy!

6. Conclusion

The two big conclusions of this paper are (1) that concepts can be learned and that this is a good thing since (2) learning models are absolutely crucial to understanding concept acquisition. We have sketched a variety of different general models for learning concepts—from hypothesis testing, perceptual learning, communication-based learning, and automatic associative learning, to the formation of experientially grounded syndrome-based sustaining mechanisms. All of these are immune to Fodor's anti-learning argument. Far from being impossible, concept learning is ubiquitous.

Given our defense of learning, it might be supposed that we have provided an argument on behalf of concept empiricism. This inference, however, would be unwarranted. While it is true that empiricists assign a central role to learning, nothing in what we have said undermines concept nativism. Indeed, we ourselves are concept nativists—just not Fodorian radical concept nativists. Nativism about concepts is often conflated with Fodorian concept nativism, but his extreme view isn't at all representative of the nativist approach. Concept nativists typically hold, as we do, that a significant number of concepts are innate.[21] But nativists and empiricists disagree not just about which concepts (if any) are innate. Importantly, they also disagree over the nature of the cognitive mechanisms involved in learning those concepts that are not innate—which both camps agree comprise the vast majority of concepts. Empiricists appeal to a small number of domain-general learning mechanisms and explain the differentiation that is found in the adult conceptual system in terms of differences that occur in experience. Nativists, by contrast, appeal to a large number of domain-specific learning mechanisms and see the differentiation as reflecting inherent features of the mind. Though the distinctive character of nativist and empiricists accounts of concept learning differ, for nativists, just as for empiricists, learning is absolutely central to the explanation of concept acquisition. The burden of this paper has been to show that the commitment to learning that both sides share is perfectly cogent. What remains to be seen is which general approach to learning will prove more successful as we come to have a deeper understanding of the mind.

References

Boyer, P. 2001: *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books.

Davidoff, J., Davies, I. & Roberson, D. 1999: Colour categories of a stone-age tribe. *Nature*, 398, 203–204.

Davidson, D. 1969: The individuation of events. In his *Essays on Actions and Events*. Oxford: Oxford University Press.

Fodor, J. 1975: *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. 1981: The present status of the innateness controversy. In his *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Fodor, J. 1990: *A Theory of Content, and Other Essays*. Cambridge, MA: MIT Press.

[21] Likely candidates for innate concepts include concepts associated with objects, causality, space, time, and number, concepts associated with goals, functions, agency, and meta-cognitive thinking, basic logical concepts, concepts associated with movement, direction, events, and manner of change, and concepts associated with predators, prey, food, danger, sex, kinship, status, dominance, norms, and morality.

Fodor, J. 1998: *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Fodor, J. 2008: *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.

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

Gallistel, C.R. 2000: The replacement of general-purpose learning models with adaptively specialized learning modules. In M. Gazzaniga (ed.) *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press.

Laurence, S. & Margolis, E. 2002: Radical concept nativism. *Cognition*, 86, 25-55.

Laurence, S. & Margolis, E. in prep: The limits of thought. *manuscript*.

Margolis, E. 1998: How to acquire a concept. *Mind & Language*, 13.3, 347-69.

Margolis, E. & Laurence, S. 2007: The ontology of concepts—abstract objects or mental representations? *Noûs*, 41.4, 561–593.

Pinker, S. 2007: *The Stuff of Thought*. New York: Viking.

Prinz, J. 2004: *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

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

Sterelny, K. 1989: Fodor's nativism. *Philosophical Studies*, 55, 119-141.

Symposium on Solutions to Fodor's Puzzle of Concept Acquisition, Annual Cognitive Science Society (2005), Stressa, Italy. <http://www.wjh.harvard.edu/~lds/pdfs/Niyogi_Snedeker-2005.pdf>

Weiskopf, D. 2008: The origins of concepts. *Philosophical Studies*, 140, 359-384.