

REVIEW ESSAY

Jerry A. Fodor, *Concepts: Where Cognitive Science Went Wrong*, Oxford University Press, 1998, ix + 174 pp., USD 55.00.

For such a short book (165 pages of text, plus a three page Preface), Jerry Fodor's *Concepts: Where Cognitive Science Went Wrong* covers a lot of ground. There are very many trees, or maybe better, multiple woods, to keep track of: theses, preliminaries, assumptions, caveats, appendices, etc. We'll start off, then, by sketching the central flow of argument, as background. We will then critically discuss two novel aspects of the book.

1. THE FLOW OF ARGUMENT

*Concepts* has three parts. Part I presents, as background, the version of the Representational Theory of Mind (RTM) that Fodor will be assuming (Chapter 1); and it presents five constraints on any acceptable theory of concepts (Chapter 2). Here, in a highly condensed form, are the results:

1.1. *Fodor's Background Theory*

A. Elements of Classical RTM

- (i) There are intentional laws
- (ii) Mental representations are the basic bearers of content
- (iii) Thinking is computation

B. Elements Added to Classical RTM by Fodor

- (iv) Meaning is information
- (v) What distinguishes co-extensive concepts is in the head

1.2. *Fodor's Constraints*

- (i) Concepts are mental particulars
- (ii) Concepts are categories (i.e., things fall under them)
- (iii) Concepts are compositional (they are constituents of thoughts)



- (iv) Quite a lot of concepts are learned
- (v) Concepts are public (i.e., they can be shared)

In Part II, we arrive at the book's central claim: the theory of concepts assumed in most of cognitive science "is in a certain way seriously mistaken" (vii). The mistake will come as no surprise to old hands: cognitive science, says Fodor, should adopt a theory of concepts which is atomistic: "... satisfying the metaphysically necessary conditions for having one concept *never* requires satisfying the metaphysically necessary conditions for having any other concept" (14). Failure to appreciate this principle is the "holistic mistake". Fodor evaluates several existing theories of concepts within cognitive science: (a) the view that most lexical items have definitions (in natural language words), where concepts, in these definitional cases, are the complex Mentalese correlates of the definition-in-words; (b) the view that concepts are prototypes; and (c) the view that concepts are abstractions from belief systems. Each of these is, in its own way, holistic. And each, Fodor argues, is importantly inadequate.

Many of the complaints that Fodor registers in this middle part are familiar: Most lexical items are undefinable; cognitive psychological research on lexical processing weighs against treating the purported definiendum as associated with a complex Mentalese expression; prototypes are non-compositional; the absence of an analytic-synthetic divide poses problems for individuating concepts as items in a web of belief; and so on. That these considerations are now well-known is no surprise, given that Fodor has been writing on these topics, not always in such a direct way, since Fodor 1970. (See Fodor 1970, 1981, 1998; Fodor and Lepore 1992, 1994, 1998, 1999; and Fodor and Pylyshyn 1988.) But, even where not novel, it is useful to have the considerations brought together into a larger whole; moreover, there is also some new material in this second part.

Part III describes "the atomist alternative", and considers in particular the implications of Fodor's own atomistic theory of concepts for concept acquisition and ontology. This last part is the most radical, and is worth the price of admission all on its own. Chapter 6, the first half of Part III, is primarily concerned with "the doorknob/DOORKNOB problem": why should it typically be experiences of doorknobs, and not (say) experiences of kosher dill pickles, which give people the concept DOORKNOB? This may sound outrageous. After all, what else would give one the concept?<sup>1</sup> But taken in the context of the view presented in Part I, it is actually quite pressing, as we will shortly see.

Having noted the general shape of the book, most of the rest of our time will be spent on two aspects of this work. Because it is impossible

adequately to comment on everything in the book, we offer no comments on Part II, much of which is familiar, concentrating instead on the most novel and provocative ideas Fodor presents. First we consider an intriguing puzzle, presented in Part I, that Fodor poses for “the Frege architecture”.

## 2. THE NEW FREGE PUZZLE

As it happens, sorting out just exactly what the puzzle is will take some effort. (That Fodor rephrases the Puzzle in so many different ways – there are, by our count, at least eight restatements of it between page 17 and page 21! – highlights that, as he himself recognises, it is not an easy puzzle to get one’s head around.) Once we’ve sorted it out, we will consider whether a cognitive scientist with Fregean leanings would truly be stymied by it. We will argue that a sufficiently crafty Fregean can overcome Fodor’s concern.

We first need to sort out some terminology. We hope the following glosses are true to Fodor’s intentions:

*Concept*: A concept is individuated by an ordered pair,  $\langle R, M \rangle$ , consisting of an external entity (the referent,  $R$ , of the concept) and a mode of presentation,  $M$  – denoted (MOP) for short – of that entity.<sup>2</sup>

*Modes of Presentation (MOPs)*: MOPs are those things, whatever they turn out to be, which differentiate between concepts with the same referent.

*Senses*: Senses are abstract entities which, in the Frege architecture, serve as MOPs. (I.e., senses, says Fodor, are Frege’s candidate for the  $M$ -element of concepts.)

*Grasping/Entertaining (a MOP)*: Grasping or entertaining is using a MOP to “present” something to thought: whatever the MOP is a mode of presentation of.

Now for the puzzle. Recall, first off, the two elements added to Classical RTM by Fodor: (B.iv) meaning is information and (B.v) what distinguishes co-extensive concepts is in the head. It is on this second point that Fodor’s version of RTM parts company with the standard Frege architecture. Fodor attempts to establish this second point by showing that, *pace* Frege and Fregeans, whatever distinguishes co-extensive concepts (MOPs) cannot be abstract, hence they cannot be Fregean senses. And if they are not abstract, it seems they must be mental; specifically, they must be syntactically individuated mental particulars. Fodor presents his reason for thinking MOPs

cannot be abstract by posing a question (in several forms) to which he thinks the Fregean cannot respond.

- [1] “Your having  $n$  MOPs for water explains why you have  $n$  ways of thinking about water *only on the assumption that there is exactly one way to grasp each MOP*. The question thus arises, what, if anything, is supposed to legitimize this assumption. As far as I can tell, unless you’re prepared to give up [the idea that MOPs are abstract], the only answer a Fregean theory allows you is: sheer stipulation” (17).
- [2] “So, the question I’m wanting to commend to you is: what, if anything, supports the prohibition against proliferating ways of grasping MOPs?” (19).
- [3] “. . . that one has as many ways of thinking of a referent as one has concepts of the referent depends on there being just one way to entertain each concept. What, besides stipulation, guarantees this?” (18).
- [4] “What guarantees that each sense can serve in only one way to present an object to a thought? I think that, on the Frege architecture [when read as including the tenet that MOPs are abstract], nothing prevents this except brute stipulation” (19).

It might seem that these are different questions: some about concepts, some about MOPs, some about senses; some about ways of thinking about a referent, some about ways of presenting an object to thought. But, as we will see, [1]–[4] are essentially paraphrases of one another.

Notice that, holding the referent element  $R$  of a concept  $\langle R, M \rangle$  fixed, concepts are individuated by the  $M$ -element alone – i.e., by MOPs alone. Notice too that ‘grasp’ and ‘entertain’ are synonyms.<sup>3</sup> Now, [1] and [2] ask, in effect: “What ensures that there’s only one way to *grasp a MOP*?” while [3] asks, “What ensures that there’s only one way to *entertain a concept*?” So, holding the  $R$  element of  $\langle R, M \rangle$  constant, [1], [2] and [3] actually ask the very same question. In particular, what each of [1]–[3] amounts to is:

*The New Frege Puzzle:* What ensures that, given an agent  $A$  who can grasp/entertain exactly  $n$  concepts (i.e.,  $n$  pairs  $\langle R, M \rangle$ ),  $A$  will have exactly  $n$  ways of thinking about  $R$ ?

And because senses are supposed to be Frege’s version of MOPs, [4] is simply the specifically Fregean version of the New Puzzle.

One more bit of terminology. We need to know what “ways of thinking (about an object)” and “way(s) to present (an object) to thought” might be. (Note that the former occurs in [1] and [3], the latter in [4]. So far as we can tell, Fodor uses them interchangeably.) Here, significant reconstruction is required, since Fodor himself uses these phrases without really defining them. That said, here’s our best guess as to what he has in mind. Ways of

thinking are to be understood as individuated according to their (normal) causal roles:<sup>4</sup>  $T_1$  and  $T_2$  are the same way of thinking about Cicero, for instance, only if the causes and effects of  $T_1$  on an agent's mental processes are the same as the causes and effects of  $T_2$  (on the same agent, of course).

To give an example, suppose that Juan is fluent in both English and Spanish, and that he can "think in either." (We realize that Fodor will consider this loose talk, since for him all thought really occurs in Mentalese. Oh well.) Suppose further that, due to his rabid Uruguayan nationalism, whether Juan is thinking in English or Spanish can actually have very subtle effects upon his thought patterns. Curiously, for instance, he is more prone to infer that Latin philosophers were gifted if he tokens 'Cicerón vivía en Roma' than if he tokens (what even he will concede is the synonymous) 'Cicero lived in Rome'. Given that these tokenings have different effects, 'Cicerón vivía en Roma' and 'Cicero lived in Rome' are ipso facto different ways of thinking, for Juan anyway. Let us further suppose that (for reasons Juan does not wish to reveal) what really makes the difference in his thought process is whether he thinks 'Cicero' or 'Cicerón'. Well then, Juan has two ways of thinking about the man Cicero.<sup>5</sup>

So here, at last, is what the New Puzzle amounts to: there ought to be, Fodor supposes, precisely as many ways of thinking about  $R$  as there are concepts *containing*  $R$ .<sup>6</sup> That is, for each pair  $\langle R, M \rangle$ , there should be exactly one corresponding causal role. But suppose, as Fodor's Frege does, that MOPs are abstract entities. Then there could well be several different causal roles for a single pair  $\langle R, M \rangle$ , because there could be different uses to which  $\langle R, M \rangle$  is put, different intentional relations that an agent could hold to a single  $\langle R, M \rangle$ , or different "vehicles" that the agent uses to think  $\langle R, M \rangle$ . Indeed, Juan provides a case in point: Frege surely isn't going to deny that 'Cicero' and 'Cicerón' have the same sense, even for Juan. But, as we've described the case, these terms occupy subtly distinct causal roles in Juan's mental life. So, Juan has one concept (i.e., the pair whose first element is the man Cicero, and whose second element is the sense shared by 'Cicero' and 'Cicerón'), but he has, associated with this single concept, two ways of thinking about Cicero: in Spanish and in English.<sup>7</sup> Having (we hope) clarified what the New Frege Puzzle is, we now consider how a Fregean might respond.

The response has two parts. First off, Fodor is saying that there is some fact which Frege cannot explain; specifically, Frege cannot explain why there is just one way of thinking for each sense. The thing is, the supposed "fact" which Frege cannot explain is no fact at all! As Frege himself saw, it just is not the case that every psychological difference (every difference in causal role, say) is accounted for in terms of a difference in sense. Hence

there is no reason why the Fregean cognitive scientist should attempt to do so. Indeed, Frege repeatedly and explicitly appeals to things like tone, coloring, and so forth – which, in agreement with Fodor, he took to be mental – as elements which may alter the mental impact of a sentence, even where there is no difference in sense: “But we must not forget that language does not simply express thoughts; it also imparts a certain tone or coloring to them. And this can be different even where the thought is the same” (Frege 1906: 295). Thus Fregeans might well reply to the question, “What guarantees that each sense can serve in only one way to present an object to a thought?”, that, as far as they can see, nothing guarantees this; adding, however, that this is fine by them, since it isn’t *true* that each *Gedanke* corresponds to just one kind of thinking. Frege elsewhere (1897: 243) makes a related point:

There is a difficulty here in that we think in some language or other and that grammar [...] is a mixture of the logical and the psychological. If this were not so, all languages would necessarily have the same grammar. It is true that we can express the same thought in different languages; but the psychological trappings, the clothing of the thought, will often be different.

Put otherwise, Fodor is just wrong to identify his MOPs with Frege’s senses. What Fodor calls MOPs, Frege would have called an amalgam of sense on the one hand, and tone on the other.<sup>8</sup> The former is abstract, the latter is mental. Moreover, Frege would surely have insisted that any attempt to tie senses to “ways of thinking” (understood as causal roles within mental processes) was hopelessly confused: Frege was not the enemy of psychology that he was sometimes made out to be, but he was clearly an enemy of psychologism. And making the nature of *Gedanke* – which are, after all, constituted by senses – depend upon one’s results in psychology is the very heart of psychologism. Granted, this may concede to Fodor what he most cares about: that there be something (a) in the head that (b) distinguishes coextensive concepts (again, using ‘concept’ in Fodor’s way, not Frege’s). But it also concedes to Frege what he most cares about, viz. that *senses* are not psychological.

The first response, then, is that the purported fact which Frege has “failed to explain” simply isn’t a fact. That is, it’s quite true that Frege does not explain why there is only one (Fodor style) “way of thinking” for each sense. But he needs no explanation for this, since it isn’t true. What Frege needs instead is an explanation of why a single sense *doesn’t* always yield the same psychological effects. But he has a ready explanation for this: namely, the psychological impact of tone and coloring.

Fodor might think that a variant of his worry can nevertheless be mounted, even granting the above. The natural worry is that there will

be different ways of thinking that come not from a difference in tone or coloring, but from a different way in which the abstract sense is grasped. If there are any cases like this, Frege will have no explanation of *them*.

This takes us to the second response on behalf of the Fregean. Notice first that it is not clear that Frege himself took senses precisely to be abstract *objects*; nor did he suppose, it now seems, that grasping a sense could be construed as very like Fodor's example, on p. 18 of *Concepts*, of imaging a diagram and using it to think about triangles. True enough, for the Fregean senses are abstract rather than mental. But they shouldn't be understood as abstract objects, of the sort numbers are. After all, one of the key roles for senses, in Frege's system, is to account for how human beings are able to apprehend abstract objects of the sort that numbers are: one apprehends a number by understanding – i.e., grasping the sense of – its numeral. So, on pain of regress, (“But how are *senses* grasped?”), it won't do to have senses being just the sort of thing that numbers are, with grasping a sense being just like apprehending a number. We characterise this difference by referring to senses as abstract entities, distinct from abstract objects like numbers that must be grasped via something else. Thus, the Fregean may escape Fodor's net at the first step, in rejecting the idea that grasping a sense is grasping a number-like object. Here again, this doesn't exactly amount to conceding Fodor's point, however, because senses, being abstract entities, will still not be in the head.

This will reasonably provoke the question, “How *do* we grasp senses, if not in the way that we grasp numbers?” And the question is pressing for the Fregean, since the abstractness of senses still seems to permit a multiplicity of ways of thinking about the referent of a concept; that is, nothing seems to tie a sense and reference to a unique way of thinking about the referent. Here's a response, deriving from Michael Dummett, which we find plausible enough, and which we take to be thoroughly neo-Fregean in spirit. One grasps a sense by understanding a sentence which expresses that sense. For instance, one grasps the *Gedanke* that  $2 + 2 = 4$  by understanding the sentence ‘ $2 + 2 = 4$ ’. On this view, there is exactly one way, that is one *method*, to grasp a sense, not numerous ways, as Fodor supposes: grasping a sense is always a matter of knowing the meaning of an expression. Moreover, meanings aren't mental, so the Fregean view doesn't collapse into Fodor's. But meanings aren't like numbers either, so the issue of “ways of grasping them” evaporates.

Now one might well ask: What is it to understand a sentence, such that doing so doesn't itself amount, in a viciously circular way, to “mentally seizing” an abstract object? Here again, the neo-Fregean may follow Dummett: to understand a sentence is to have the capacity to use the sentence

correctly, in accordance with the conventions of the language. That is, knowing an expression's meaning amounts to having a certain linguistic ability. It does not involve standing in some queer relationship to some number-like thingy. Of course, having noted this Dummett-inspired way out for the Fregean, it must be granted that it brings with it at least a soupçon of holism. For one thing, on this Dummettian view, it is whole sentences that get used; and, generally speaking, a sentence contains more than a single word. But, most importantly, it's part and parcel of this Fregean view that grasping the sense of a single word is a matter of seeing the word's contribution to a whole series of sentences.<sup>9</sup> (One cannot find a pattern of meaning-contribution in a single sentence!) Hence this neo-Fregean approach requires there to be a whole language, with conventions governing sentence formation and use. ("To understand a sentence means to understand a language. To understand a language means to be master of a technique." *Philosophical Investigations* 199.) Fodor may complain about this holism, but it's not a problem for the Frege architecture that it may not permit the sort of radical atomism that Fodor desires. Why should the Fregean worry that it is holism which, in part, allows her to respond to Fodor's objections?

But now it may be objected that, while the Dummettian view we are endorsing does admit only one *method* of grasping a sense, this does not yet entail that there is only one *way* of grasping a given sense. In fact, there very clearly remain numerous "ways" within the understanding-a-sentence "method": one for each sentence that expresses the sense. To repeat, grasping a sense is always a matter of understanding a sentence which expresses that sense – but *any* sentence will do. And, in general there will be many such sentences, hence many ways of grasping a sense, one (at least) for each sentence that expresses the sense. As we said above, our example with Juan is just such a case. 'Cicero' and 'Cicerón' express the same sense, nonetheless they are the vehicles for two distinct ways of thinking about Cicero. And here we have precisely one method for grasping, i.e., understanding a name, but two ways of thinking. However, the answer to this can be easily imagined: these differences in "ways of grasping" all fall within the rubric of tone/coloring differences within the sentences. The general objection, recall, was that there might be cases of one-sense, many "ways of thinking" which arise not from tone/coloring, but from different ways of grasping senses – cases which, therefore, Fregeans could not explain. But it now turns out that the various "ways" all involve distinct sentences. Sentences which, it's safe to say, differ in tone/coloring. So the existence of such "ways" is of no concern to the Fregean.



In sum, Fregeans can reject the idea that grasping a sense amounts to standing in some relation (one of many possible ones) to a number-like object; and, they can equally reject the idea that senses must, all on their own, account for every possible difference in Fodor style “ways of thinking”. Making these two moves, which are both thoroughly in the spirit of Frege, the Fregean cognitive scientist can reply to Fodor’s puzzle.

We turn now to a second aspect of the book, namely Fodor’s new account of concept acquisition. On a first reading, this account is nothing short of shocking, especially coming from Fodor. It seems to undermine, if not downright contradict, much of what Fodor has been espousing throughout his career. A closer reflection, however, reveals that Fodor’s overall position is still very much intact. In fact, this new position is an attempt to reconcile tensions between the various views Fodor wants to hold, and can be seen as a continuation of the problem Fodor sets himself in *The Elm and the Expert* (Fodor 1994). We argue that Fodor is unsuccessful in meeting all of the constraints imposed on his view by holding these various positions. Our point is not that Fodor has not yet got the details right – that is to be expected with a new position – but that it seems unlikely any position can mutually satisfy all of the constraints he has imposed.

### 3. FODOR’S NEW ACCOUNT OF CONCEPT ACQUISITION

About twenty-five years ago now there was an argument that went like this: concept learning is a process of hypothesis formation and confirmation. Hypotheses need a medium in which to be formulated, and confirmation is a computational process within this medium; hence, any creature that engages in concept acquisition must antecedently possess an internal representational system: a language of thought. This argument was stated in, for example, *The Language of Thought*, a book by a certain J. A. Fodor. He there writes:

We have been considering some of the ways in which viewing the concept learning task as essentially involving inductive extrapolation commits one to postulating a representational system in which the relevant inductions are carried through. I think it is worth emphasizing that no alternative view of concept learning has ever been proposed . . . . (Fodor 1975, p. 41)

In *Concepts*, Fodor seems to be offering an alternative view. The central idea is that concept acquisition, rather than being a matter of hypothesis formation and testing, is just a matter of coming to be “nomologically locked to the property that the concept expresses” (p. 125). Concept acquisition is not, then, an inductive learning process: “If you assume... the

locking model of concept acquisition..., then you can't assume that hypothesis testing is an ingredient in concept acquisition" (p. 126). In this section we will consider whether, when all is said and done, Fodor really does provide a viable alternative view of concept acquisition.

Here is the game plan. We will begin by rehearsing *Concept's* account of concept *possession*. We will then explain a series of constraints, explicitly endorsed by Fodor, that any theory of concept *acquisition* must meet. Having done that, we will consider whether Fodor can offer an account of concept acquisition that meets all of the constraints. We argue that, probably, he cannot.

First up, then, concept possession. According to Fodor, an agent possesses a concept, say SUSHI, if the agent is locked to the property which SUSHI expresses. Notice that the relata in the locking relation are properties and agents – or, as Fodor sometimes has it, properties and “minds”. (He also sometimes uses the term ‘resonates to’ for this person::property relationship. We take ‘resonates to’ to be a synonym of ‘locks to’.) That these are the relata may suggest, to the incautious reader, that the only things which really bear content, on Fodor's new view, are persons, as contrasted with sub-personal systems or agents.<sup>10</sup> This reading can equally be taken away from claims like: We have a concept expressing some property when things having the property strike *us*, i.e., whole persons, in a certain way (p. 136). But this can't be what Fodor really has in mind. After all, Fodor in Part I explicitly laid down as part of his background theory that mental representations are the basic bearers of content. And too, as if the former consideration were not enough, how can property-locked *whole-minds* do the job required by RTMs? If nothing sub-personal possesses content, then none of the causal interactions between sub-personal systems can be viewed as computational – because they are not causal relations among mental *representations*. Something gets to be such a representation only if it is a syntactically structured item, where that means, in part, that its content is exhaustively determined by the contents of its syntactic constituents and their relations, etc. But then forms lacking intentional content, i.e., *mere* shapes, cannot have syntax. (This point will play a pivotal role at the end of the paper.) Put otherwise, not every formal description is a syntactic description; only contentful things, whose content is determined compositionally, have syntax. So if the locking model entailed that nothing sub-personal possesses content, then in invoking the locking model, Fodor would undermine the RTM itself. The conclusion from the above considerations must surely be that, though Fodor doesn't himself stress the point, on the locking model, sub-personal items do possess content. Since

he does not specify what these might be, we simply refer to them as ‘neural structures’.<sup>11</sup> More on this shortly.

That, in barest outline, is the locking model of concept possession. It might seem obvious what story ought to be told about concept acquisition, given the foregoing story about concept possession. To acquire a concept, say *SUSHI*, is simply for the agent to *become* locked to the property *sushi-hood*. To account for the point about sub-personal entities inside the agent having content, we might introduce the notion of “binding,” which is the sub-personal version of locking. Binding obtains between a neural structure and a property. Given this, to acquire a concept would just amount to some neural structure within the person’s brain becoming bound to a property. (And, of course, once that happens, the agent is then locked to the property as well, in virtue of her neural structure being bound to it.)

What we want to consider next is whether the locking model, as we have reconstructed it, really can sustain a non-inductive view of concept acquisition. Specifically, any account of concept acquisition, says Fodor, needs to meet a host of constraints. And we are not convinced that this model can meet all of them. Here, then, are the constraints.

### 3.1.

- (a) The theory of concept acquisition should not entail radical nativism about concept possession;
- (b) It must distinguish merely having an experience from actually tokening a concept;
- (c) It cannot end up being an inductive story;
- (d) It must nevertheless share the virtues of the inductive account, including especially overcoming the doorknob/DOORKNOB problem;
- (e) It must be atomistic.

Some of these could use explaining. So, with the end both of raising our objection, and also of articulating several of the central ideas of the book, we will spend some time spelling them out – and noting how, at least at first glance, Fodor’s locking view seems to meet them.

To repeat, on Fodor’s current view, “acquiring a concept is getting *nomologically locked* to the property that the concept expresses” (125). What’s nice about this locking story is that, *prima facie*, it takes much of the bite out of concept nativism: while neurologically speaking all mental representation forms still have to be innate – put roughly, pretending that the “neural structures” can be modeled by binary code, every usable string of ones and zeroes must be innately available – semantically speaking, the content of those strings can be endowed by experience. As, indeed, they must be, at least in the early stages of concept acquisition. For instance, while Plato

may have had the ability to token the form ‘1100111001’, which now means *carburetor*, he didn’t exactly have the concept CARBURETOR, because his string wasn’t locked to carburetors – combustion engines being a bit thin on the ground at the time. This is constraint (a). Moreover, and in accordance with constraint (c), on the locking view, “learning” this content doesn’t require infants to frame hypotheses of the form ‘1100111001 must mean carburetor’ in which the concept itself is used in the statement of the hypothesis, thus making its content (and not just its form) seem innate. Instead, the string acquires its content through brute causal interaction with the environment.

So, one replaces innate contents with innate forms, plus innate mechanisms for pairing forms with external entities. The result is that, in one (non-worrisome) sense, the concept CARBURETOR is innate – as are, given atomism, most lexical concepts. That’s because the neural structure, e.g., the string of ones and zeroes, is innate; and so is the mechanism for locking it to carburetors, should they appear in the environment. But in another sense – the one which so exercises conceptual Empiricists – CARBURETOR and other lexical concepts are acquired through experience, since that’s how the string in question comes to be locked to its content-giving property. So called mad dog nativism made bland, if you will. (Fodor concludes the chapter, on p. 143, saying: “Maybe there aren’t any innate *ideas* after all”. Honest. We’re not making this up.)

As for constraint (b), Fodor needs it because, as we will explain shortly, he wants to say that there are appearance properties (e.g., *doorknobhood*) which are not sensory properties (e.g., *red*). But then there must be some difference between them. The distinction Fodor proposes is “*S* is a sensory property only if it is possible to have an experience of which *S*-ness is the intentional object (e.g., an experience (as) of red) *even though one hasn’t got the concept S*” (p. 135, footnote 10, emphasis in original). That is, a “red-experience” is not a tokening of the RED concept – though, of course, it could lead to a tokening of the RED concept. To have an experience of *redness*, it is sufficient to have sensory organs that can produce the experience. But tokening a concept requires having a thought of the right kind. For example, tokening the concept RED requires having a thought like *that’s red* (p. 135, footnote 10). This is what constraint (b) says. (Put a more traditional way: in the case of sensory properties, applying the concept is something over and above experiencing the extension of the concept.)

Constraint (d) requires an account of concept acquisition to explain why it is, typically, encounters with doorknobs in virtue of which we get locked to DOORKNOB. Fodor himself has notably claimed that,

... one of the distinguishing characteristics of concept learning is the *nonarbitrariness* of the relation between what is learned and the character of the experiences that occasion the learning. ... To put it mildly, it seems unlikely that any theory radically incompatible with ... [an inductive account of concept learning] ... could account for the nonarbitrariness of the relation between what is learned and the experiences that occasion the learning (Fodor 1975, pp. 37–8, emphasis in original).

This constraint isn't that easy to get one's head around, so here is another way of putting it: If getting endowed with content really is a brute causal process, a process of coming to be nomologically locked by something other than induction, why can't '1100111001' resonate to carburetors as the result of encountering a bad batch of Steak Tartar? Why is experience of carburetors so central? To see the seriousness of this question, compare the following: in an important sense, human antibodies are "outward looking," in that they map onto specific kinds of bacteria out in the environment. Hence, albeit in a quite different sense, an internal antibody may be "locked" to an external bacterium. Now, one way to get an antibody which is locked to a specific bacterium is to encounter the bacterium, internalise it, and develop an immune response. This is comparable to attaining a concept by experiencing its associated property. But encountering the bacterium itself is not the only way to get an antibody for it. Another quite common way is to get the antibody from one's mother, through breast milk. (We're told that's how one typically gets Immunoglobulin G.) These days, of course, one can even get artificial antibodies, through an injection of a lab-manufactured compound. So, why does "antibody-bacteria locking" in humans seem so different from concept-property locking, if both are brute causal? In particular, why *can't* you get CARBURETOR, content and all, from breast milk, or from an injection? This is the doorknob/DOORKNOB problem. (Though the antibody-concept comparison is ours, we doubt Fodor would disapprove of it.) Constraint (d) says: any adequate theory of concept acquisition must have a solution to it.

Now, Fodor spends a considerable amount of time trying to meet this latter constraint. Being a doorknob, he says, is being the sort of thing stereotypical instances of which lead normal humans to acquire the concept DOORKNOB. He writes:

... *being a doorknob* is having that property that minds like ours come to resonate to in consequence of relevant experience with stereotypic doorknobs. That, and *not* being learned inductively, is what explains the content relation between DOORKNOB and the kinds of experience that typically mediates its acquisition. (p. 137, emphasis in original)

It is this metaphysical fact – to be a doorknob is to interact in the right way with human psychology – which explains why doorknob-experiences are so central to the acquisition of DOORKNOB: being a doorknob is

cached in terms of causing the acquisition of DOORKNOB, so of course experience of doorknobs ends up being central. What's shocking about this, *prima facie*, is that whether something is a doorknob (or sushi, or a miniskirt) is, on Fodor's current story, constituted by whether we take it to be one. Fodor goes postmodern? That truly would be alarming. But the situation isn't quite so extreme. First off, as Fodor stresses, there are minds which can and do acquire DOORKNOB, so there really are doorknobs. Besides, it's not as if *everything* is like doorknobs: gold, atoms, viruses, etc. Indeed, one key burden of Chapter 7 is to argue that natural kind concepts are importantly unlike DOORKNOB (and RED), in that they are individuated not in terms of likeness of effects *on us*, but in terms of their effects on other things. So some properties – notably, the ones investigated by the sciences – aren't mind-dependent appearance properties in the way that *doorknobhood* is.

The problem is, Fodor concentrates so much attention on the doorknob/DOORKNOB constraint that he seems to forget to fill in the rest of the story of concept acquisition. Worse, once one tries to flesh out the bare claim that neural structures come to be bound to properties, it becomes clear that Fodor has too many constraints in play. Or so we'll now argue.

To begin with, a worry about constraint (c). We already complained that Fodor doesn't say much about how acquisition occurs. But the little he does say invites doubt that he really has proposed an *alternative* to concept learning as an inductive process. For instance, in attempting to explain why we learn the concept *X* and not the *X* stereotype from stereotypic examples of *X*, Fodor says "it's a law about our kinds of minds that they are set up to make *inductions* from samples consisting largely of stereotypic English sentences to the concept ENGLISH SENTENCE . . . and not . . . to the concept STEREOTYPIC ENGLISH SENTENCE" (p. 139, our emphasis). He also says that "being a doorknob is having the property to which minds like ours *generalize* from experiences (as of) the properties by which the doorknob stereotype is constituted" (p. 140, our emphasis). To the extent that Fodor has any account at all about the "resonating to" process, it seems to be inductive after all.

So even the little Fodor does say about concept acquisition suggests that it may fail to meet one of the constraints. But things get worse. Even putting this aside, it turns out that each attempt to flesh out concept acquisition fails to meet at least one of the constraints in 3.1. Our conclusion will be not that we are insufficiently inventive, but that Fodor has placed too many constraints on himself. And that some of them ought to go.

First attempt. One natural thought is that, for example, instances of red trigger the concept RED. And, in general, locking comes into being because the property in the environment triggers the concept in the agent's brain. Fodor quite rightly does not endorse this view, however, for while it would be brute causal it has two key flaws. First, if instances of redness trigger the concept RED then it is not possible to have red experiences without having the concept RED, so Fodor's distinction between sensory concepts and appearance concepts, enshrined in constraint (b), collapses. Second, and more importantly, if causal interactions with red things in the world trigger the concept RED, then we innately possess *a disposition for red things to strike as red*; i.e., on Fodor's current view we innately possess the concept RED. And we would also innately possess DOORKNOB, SUSHI, and the rest if they too are triggered by a single experience. So constraint (a) is not met.

Second attempt. Maybe the only problem with the triggering view is that it posits locking after a single experience of the relevant property. Better might be the following: the agent several times encounters a doorknob, and applies the concept DOORKNOB each time. After enough iterations, a pattern emerges, of doorknobs causing the agent to token DOORKNOB. At this point, locking between the property *doorknobhood* and the agent has occurred. But let's slow down a bit here, and consider the first encounter with a doorknob. Surely the agent cannot apply the concept DOORKNOB to the thing experienced, at the moment of that first encounter! She isn't locked to the property yet. Which means, on the locking view, that she doesn't have DOORKNOB yet. (Unless, of course, DOORKNOB is innate. Which, according to constraint (a), it cannot be.) Well but, if this problem arises with respect to the first encounter with doorknobs ... why is the agent any better off the second time? She still doesn't have the concept. So she still can't apply it. In which case, she can't establish a pattern of experiencing-doorknobs-and-subsequently-applying-DOORKNOB.

Third attempt. The problem with the first story was that locking was established from the very first encounter. This made the concept innate, and it blurred the distinction between sensory properties and (what Fodor calls) other appearance properties. The problem with the second story was that concepts were being employed before the requisite property-locking had actually occurred, which put the cart in rather the wrong spot vis-à-vis the horse. One can avoid both of these failings by supposing that: (i) there are repeated occurrences of the property causing a neural structure to be activated, but such that (ii) the neural structure in question is, as yet, without content. Surely that is how locking gets set up.

But, on second thought, thinking is computation (so says the background theory, described in Part 1 of the paper), and concepts are compositional; i.e., they are constituents of thoughts (this, recall, was one of Fodor's constraints, also laid out in Part 1). So, concepts are the sorts of things which have not just a shape, but a syntax. But, as we noted above, when discussing what we called binding, forms must have content, if they are to have syntax. Hence the neural structures, if they don't already have content, don't have the appropriate sort of form either, namely, syntactic form. That is, it isn't enough to have any old internal thingy taking part in a "nomic dance" with some external property. (Thermostats are, of course, a case in point: they are indicators, but they don't have any syntactic states.) Rather, in the brain case, the neural structure must have certain other features, not least of which is having a syntax. And this it can get only when wholes get their content determined on the basis of the contents of their constituents.

A natural reply is that one surely can introduce a syntactic item into a language, without having to give it any content. Indeed, this is routine in formal languages: one introduces a predicate, for instance, without giving its interpretation. So it can't be true that having syntax requires having semantics. We agree. Assuming that there exists a whole system of syntactic forms already, with rules governing their formation and their compositional interpretation, it makes eminent sense to speak of an uninterpreted item having syntax. One could even have such an uninterpreted syntactic item becoming bound to an external property over a series of property-causes-structure iterations. And, in such a scenario, one would truly have the acquisition of a new *concept*. But now, if *this* is the story one tells about concept acquisition by humans, what is being violated is constraint (e), atomism – the most important constraint of all for Fodor. Put in Fodor's terms, it just wouldn't be true that "... satisfying the metaphysically necessary conditions for having one concept *never* requires satisfying the metaphysically necessary conditions for having any other concept" (14).<sup>12</sup> Rather, having a concept (or anyway, acquiring a concept) would require the prior existence of a whole system of meaning-bearing items.<sup>13</sup> That's what would help make it the case that there would be syntactic structures at all, which would then allow there to be one without content, but with syntactic structure.

To give a rough and ready slogan, which encapsulates this last point: even if semantics *per se* weren't holistic, syntax surely is.<sup>14</sup> And since concepts have both syntactic form and content, concepts turn out not to be atomic. (Can one even imagine a language that had exactly one Noun Phrase, and nothing else? What would that mean? What would make it the



case that the expression was headed by a Noun? And what would make it a phrase? In fact, what would make it a syntactic item at all, rather than some sort of gesture?)

Fourth attempt. Perhaps in our reconstructions to this point we have been too hasty in trivialising the role of stereotypes. It might be that we can avoid all three previous objections and still maintain that (i) there are repeated occurrences of the property causing a neural structure to be activated, but such that (ii) the neural structure in question is, as yet, without content. We have been supposing that the neural structure in question, once endowed with content, will have the content of the concept being acquired. For example, in acquiring the concept DOORKNOB, we have been assuming that the neural structure activated by repeated encounters with stereotypic doorknobs will come to have the content *doorknobhood*. However, it might instead come to have the content *stereotypic doorknob*. Repeated encounters with stereotypic doorknobs are required to endow the relevant neural structure with content because the process is (you're not going to like this) *inductive*.<sup>15</sup> Once we learn the concept DOORKNOB STEREOTYPE (or is it DOORKNOB-STEREOTYPE?) we automatically lock to the concept DOORKNOB, because that's the kinds of minds we have. Notice that this commits Fodor to a phenomenalist semantics for concepts of the form X STEREOTYPE, as you might expect given that stereotypes are statistically defined in terms of sensory properties (the hypotheses, recall, are stated in terms of sensory concepts). But Fodor avoids a more pervasive phenomenalist semantics because concepts in general are triggered and not learned inductively, hence their content is not reducible to the vocabulary of the hypotheses in terms of which X STEREOTYPE concepts are learned.

Notice how this account apparently avoids our objection concerning syntactic holism. The idea would be that syntactic items without content can be introduced because, on this view, there does exist a whole system of syntactically structured items – which, however, are not themselves concepts. Furthermore, ostensibly radical concept nativism is avoided because the triggering of concepts by the environment is mediated by the inductive learning of the concept's stereotype, without presupposing the concept itself. Appearances can be deceiving, however. The vocabulary for the hypotheses in terms of which a stereotype concept is learned includes sensory *concepts*. But how are sensory concepts to be learned? Clearly not inductively, since there is no more primitive vocabulary to which we can appeal. Yet if they are triggered by the environment, as Fodor suggests (personal communication), the concepts turn out to be innate after all, as we argued above (first and second attempts). This by itself might be

something Fodor could live with, since it only entails the innateness of sensory concepts.

But there is a final worry. Suppose Fodor bites the bullet and accepts that sensory concepts are innate. Then allegedly this gives him the resources to account for all of concept acquisition. But recall that repeated experiences of stereotypic doorknobs enable us to inductively learn the concept DOORKNOB STEREOTYPE, on this view. Now DOORKNOB STEREOTYPE, superficially at least, has a constituent structure which includes DOORKNOB. And in fact, if it did not have a constituent structure, nothing would support inferences (inductive or otherwise) from DOORKNOB STEREOTYPE to DOORKNOB. And if there is no support for such inferences, why does DOORKNOB STEREOTYPE trigger DOORKNOB and not SUSHI? In particular, if DOORKNOB STEREOTYPE is supposed not to have a constituent structure, then the doorknob/DOORKNOB problem remains; i.e. condition (d) is violated. But if DOORKNOB STEREOTYPE does have a constituent structure, then learning it presupposes the concept DOORKNOB. All of our concepts turn out to be innate after all, and we are back to radical concept nativism, violating condition (a).

The upshot is that, once one tries to flesh out the story of concept acquisition, on the locking model of concept possession, the various attempts all fail of (at least) one of Fodor's constraints.

#### 4. CONCLUSION

*Concepts* is so densely packed with fascinating details, intriguing argumentative twists, and innovations that, though the forest is sometimes a bit elusive, the various trees/woods are nonetheless well worth the effort. That said, there are of course some aspects that we find ultimately unsatisfying. We specifically noted two. First, with respect to Part I of the book, we doubt that "the New Frege Puzzle" actually shows what Fodor thinks it does, viz. that, *pace* Frege, senses must be mental. It may show that senses cannot be exactly like numbers; and that senses cannot be expected to capture every psychological nuance that might be called a 'different way of thinking'. But this is no skin off the Fregean's nose. Secondly, with respect to Part III of *Concepts*, we remain unconvinced that Fodor has offered a viable alternative to his older, inductive, account of concept acquisition. Our view is that no position could mutually satisfy all of the constraints Fodor has imposed on an account of concept acquisition, and that something fundamental to Fodor's overall project (like say radical atomism) will have to be abandoned in order to produce a position that

meets the various concerns. But then, as Fodor himself recognizes, this is very new terrain, and two chapters in a brief book can't really be expected to produce anything like the final word.

## NOTES

<sup>1</sup> We adopt Fodor's notational conventions in what follows. In particular, names of concepts are written in capitals, and names of semantic values appear in italics. Thus 'RED' expresses *being red* is true.

<sup>2</sup> Where it makes no difference, we will speak of a concept as *being*, rather than being individuated by, such an ordered pair. This does not, we think, depart significantly from Fodor's usage. He writes in a footnote, for instance, that "a concept is a MOP together with a content; and I've taken an informational view of the individuation of contents" (20). Of course what Frege calls 'concepts' are certainly not individuated by anything like  $\langle R, M \rangle$ , i.e., a referent-sense pair. To the contrary, Frege uses the word 'concept' (*Begriffe*) not for something which *designates* a referent, but for something which *is* a referent (and is determined by a sense). Specifically, concepts, for Frege, are the (unsaturated) referents of predicates. For the details, see especially Frege (1892a, 1891), as well as Dummett (1973). That said, we will use Fodor's terminology in what follows, even where Fodor is discussing Frege.

<sup>3</sup> It might be natural to distinguish these terms by considering 'grasp' to be dispositional and 'entertain' to be episodic. However, Fodor explicitly denies such a distinction. "I use 'entertaining' and 'grasping' a MOP (/concept) interchangeably" (p. 17).

<sup>4</sup> Notice that admitting a normal causal role is unproblematic for Fodor provided meaning is not constituted thereby.

<sup>5</sup> There are many other such examples. Frege explicitly discusses 'horse' versus 'steed', 'dog' versus 'cur', 'but' versus 'and', the passive versus the active voice, and so on. (Dennett (personal communication) has suggested another nice example: 'barf' and 'vomit!') The point of all of these examples is simply that words sharing the same sense can play different causal roles in our thinking.

<sup>6</sup> See note 2 for a reminder of the terminology here. In particular, on Fodor's usage, concepts do indeed contain objects as elements.

<sup>7</sup> Here's a way of putting the point, in more properly Fregean terms: it looks like there could be a whole series of thinkings of different sorts, each having a different impact on the thinker, despite the fact that these "thinkings" were of the very same *Gedanke*. As we will see, this possibility was explicitly anticipated by Frege. By the way, it is of course true that there are two distinct ways of grasping a sense even on Frege's view. Namely, one can refer to the sense itself (i.e., one can think about the sense), or one can use the sense to present a referent (i.e., one can think with the sense). Clearly the former way of grasping is not at issue in what follows.

<sup>8</sup> For more on this distinction, see Frege (1892a, 184; 1892b, 155; and 1918, 330–331). Dummett (1973) contains a very useful eponymous chapter as well.

<sup>9</sup> The idea that word-sense (and word-reference) is grasped by first seeing a *syntactic* pattern in a *series* of sentences, and then considering the contribution of that pattern to sentence-meanings, is most clearly illustrated in grasping functions. Frege (1891, 133) writes, for example:

People ... recognize the same function again in

$$2 \cdot 1^3 + 1,$$

$$2 \cdot 4^3 + 4,$$

$$2 \cdot 5^3 + 5,$$

only with different arguments, viz. 1, 4 and 5. From this we may discern that it is the common element of those expressions that contains the essential peculiarity of a function; i.e. what is present in

$$2 \cdot x^3 + x$$

over and above the letter 'x'.

Notice: we fix on a common element *in the expressions*, and then discern what it “contains” (which is the word Frege uses when he wants to leave open whether the “thing contained” is the sense or the referent). Our point is, the process demands that we consider a series of expressions, not just the word in isolation.

<sup>10</sup> The distinction here involves whether meaning is simply ascribed to us as a whole or whether we possess internal structures that have content, in virtue of which meaning is ascribed to persons.

<sup>11</sup> It is an open question whether these “neural structures” correspond to neural types. Our point is only that certain neurological items serve as tokens of the primitive bearers of intentional content.

<sup>12</sup> Note that color experience is not atomistic in Fodor’s sense. It is not possible to experience red without also being able to experience other colors because of the constitution of our visual system. The pigments in our cones are more likely to absorb photons of certain wavelength than others, but this is merely probabilistic. And when they do absorb a photon, regardless of its wavelength, their response is always the same; thus “people with only a single type of cone are unable to experience color” (Kandel et al. 1995, 456). Color experience is constructed by the brain from the comparative responses of entire systems of cones. Thus color experience requires at least two distinct kinds of photoreceptors, so no color can be experienced unless more than one color can be experienced.

<sup>13</sup> This is much the same point Fodor makes in discussing the acquisition of natural kind concepts as such. Note also that, for this reason, it won’t do to suppose that Fodor could adopt a weaker version of atomism – in which one need not possess any *particular* concepts in order to acquire a given concept, but nonetheless one must possess some concepts. How then would we ever come to have any concepts, since *none* are innate?

<sup>14</sup> . This won’t be a grand surprise to Fodor. Indeed, Fodor and Lepore (1992: 55) once suggested that:

... the sentence is the unit of *syntax*; that is, that words couldn’t have the *syntactic* properties they do if they did not occur as constituents of sentences. This suggestion seems plausible enough; on the face of it, syntactic properties look to be the ones that words have in virtue of their relations to the sentences that contain them.

But surely the same lesson applies to mentalese “words”, i.e., concepts.

<sup>15</sup> Fodor himself (personal communication) suggested that this is the way to flesh out his account of concept acquisition. We are grateful to him for this and other feedback. Thanks also to: Dorit Bar-On, Andy Brook, Lenny Clapp, Dan Dennett, Marc Hauser, Tim Kenyon,

Paul Pietroski, and audience members at the 1999 Interamerican Philosophy Conference, Puebla, Mexico, where the paper was presented in draft form under the title “Reflections on Fodor’s *Concepts*”.

## REFERENCES

- Beaney, Michael: 1997, *The Frege Reader*, Blackwell, Oxford.
- Dummett, Michael A. E.: 1973, *Frege: Philosophy of Language*, Duckworth, London.
- Fodor, Jerry A.: 1970, ‘Three Reasons for Not Deriving “Kill” from “Cause to Die”’, *Linguistic Inquiry* **1**, 429–438.
- Fodor, Jerry A.: 1975, *The Language of Thought*, Harvard University Press, Cambridge MA.
- Fodor, Jerry A.: 1981, ‘The Present Status of the Innateness Controversy’, in J. A. Fodor, *RePresentations*, The MIT Press, Cambridge MA.
- Fodor, Jerry A.: 1994, *The Elm and the Expert*, The MIT Press, Cambridge MA.
- Fodor, Jerry A.: 1998, *In Critical Condition*, The MIT Press, Cambridge MA.
- Fodor, Jerry A. and Ernest Lepore: 1992, *Holism: A Shopper’s Guide*, Blackwell, Oxford.
- Fodor, Jerry A. and Ernest Lepore: 1994, ‘The Red Herring and the Pet Fish: Why Concepts Still Can’t be Prototypes’, *Cognition* **58**, 253–270.
- Fodor, Jerry A. and Ernest Lepore: 1998, ‘The Emptiness of the Lexicon’, *Linguistic Inquiry* **29**.
- Fodor, Jerry A. and Ernest Lepore: 1999, ‘Impossible Words’, in E. LePore and Z. Pylyshyn (eds.), *What Is Cognitive Science?*, Blackwell, Oxford.
- Fodor, Jerry A. and Zenon Pylyshyn: 1988, ‘Connectionism and Cognitive Architecture: A Critical Analysis’, *Cognition* **28**, 3–71.
- Frege, Gottlob: 1891, ‘Function and Concept’. Reprinted in Beaney (1997).
- Frege, Gottlob: 1892a, ‘On Concept and Object’. Reprinted in Beaney (1997).
- Frege, Gottlob: 1892b, ‘On *Sinn* and *Bedeutung*’. Reprinted in Beaney (1997).
- Frege, Gottlob: 1897, ‘Logic’. Reprinted in Beany (1997).
- Frege, Gottlob: 1918, ‘Thought’. Reprinted in Beany (1997).
- Kandel, E. R., J. H. Schwartz and T. M. Jessell: 1995, *Essentials of Neuroscience and Behavior*, Stamford, Connecticut: Appleton and Lange.

Robert J. Stainton  
 Department of Philosophy  
 Carleton University  
 Ottawa, Ontario  
 Canada

Christopher Viger  
 Department of Philosophy  
 Dalhousie University  
 Halifax, Nova Scotia  
 Canada