PHILOSOPHICAL PSYCHOLOGY, VOL. 12, NO. 1, 1999

Concepts: where Fodor Went Wrong

ALEX LEVINE & MARK H. BICKHARD

ABSTRACT *In keeping with other recent efforts, Fodor's CONCEPTS focuses on the metaphysics of conceptual content, bracketing such epistemological questions as, "How can we know the contents of our concepts?" Fodor's metaphysical account of concepts, called "informational atomism," stipulates that the contents of a subject's concepts are fixed by the nomological lockings between the subject and the world. After sketching Fodor's "what else?" argument in support of this view, we offer a number of related criticisms. All point to the same conclusion: Fodor is ultimately not merely bracketing the epistemology of conceptual content; his theory makes answers to the epistemological questions impossible.*

1. Introduction

Readers of Jerry Fodor's recent work, in particular his (1992) collaboration with Ernest Lepore, *Holism: a shopper's guide*, and his (1990) anthology *A theory of content and other essays*, will no doubt have suspected Fodor's latest book, *Concepts: where cognitive science went wrong*, Oxford, Oxford University Press (1998) was in the pipeline. While decrying the dangers of treating semantic properties as "anatomic," *Holism* is "officially neutral" on the theory of concepts. In their attack on arguments for semantic holism, Fodor and Lepore scrupulously avoid committing themselves to conceptual atomism, preferring a refutation of arguments for holism free of any dependence on a defense of the latter doctrine. And yet one has the sense that the stage has been set for this defense, and so its appearance in *Concepts* (1998) hardly constitutes a surprising turn.

Equally unsurprising is the fact that the basic structure of Fodor's defense is that of a "what else?" argument, a strategy familiar from Fodor's (1975) *The language of thought* and every major work since. Certain plausible, "non-negotiable" conditions are proposed as constraints on candidate theories of concepts. The theory on which concepts are definitions meets most of them, but fails to meet others. Other going theories, those which hold that concepts are prototypes or stereotypes fail on a different subset of the non-negotiable conditions. What both definition-based and stereotype-based theories of concepts have in common is that they hold

Alex Levine, Department of Philosophy, 15 University Drive, Lehigh University, Bethlehem, PA 18015, USA; e-mail: ATL2@lehigh.edu, http//guava.phil.lehigh.edu/alexhome.htm. Mark H. Bickhard, Department of Philosophy, 15 University Drive, Lehigh University, Bethlehem, PA 18015, USA; e-mail: MHB0@lehigh.edu, http://www.lehigh.edu/~mhb0/mhb0.html

0951-5089/99/010005–19 $7.00

that a concept is constituted in part by its inferential role; they entail inferential role (IR-) semantics. In light of the shortcomings of such theories, the view which takes concepts as informational atoms emerges as a plausible alternative. Conceptual atomism has not been demonstrably proved, Fodor claims, but in the absence of any other candidate capable of meeting our demands, there is a presumption in its favor.

It is worth noting that *Concepts* does not actually present an atomistic theory of concepts in any detail. What it does offer is an account of the metaphysics of conceptual content, an account which combines conceptual atomism with informational semantics. Epistemological issues, such as how we might come to *know* the contents of our concepts, are pointedly turned aside; the suggestion seems to be that they aren't worth discussing until a metaphysical foundation has been established.

In this book Fodor wields his famous wit and charm with all his customary skill (nor will fans of Fodor's Auntie be disappointed). The text originated as the 1996 Locke Lectures, and it has preserved some of the flavor of oral delivery. *Concepts* displays Fodor's well-known penchant for arguing an unpopular position from widely accepted premises. It is an engaging and important work and, we will argue, a deeply mistaken one.

After sketching the argument of Fodor's book, we will offer a line of criticism whose aim is to show, first, how heavily Fodor relies on the assumptions that to give an account of concepts is to do metaphysics, and that toward this end all matters epistemological may safely be set aside. Next, we argue that when our task is the development of a theory of concepts, this practice is unacceptable; a theory of concepts developed using Fodor's procedure risks failing to meet the demands we place on such theories. Fodor's attempt, we believe, succumbs to this risk [1].

While our discussion is formally a critical review of *Concepts*, we believe it has wider implications. In a sense, *Concepts* represents the maturation of a project on which Fodor has been embarked for years. He has drawn out, perhaps more insightfully than anyone else, the consequences of key presuppositions that lie, frequently unacknowledged, behind much of our work in cognitive science and the philosophy of mind. We believe Fodor is right that most cognitive scientists are committed, knowingly or otherwise, to something like the theory of concepts Fodor sets out in this book. If we are right about the shortcomings of that theory, its failure is thus of much broader significance than might be thought.

2. Sketch of the argument

The whole point of coming up with a theory of concepts, for Fodor, is that such a theory is needed in order to complete the Representational Theory of the Mind (RTM). Chapter 1 begins by warning the reader to expect not a defense of RTM, but rather an exploration of the consequences of taking RTM seriously. The remainder of the opening chapter is an exposition of Fodor's version of RTM, which he treats as the conjunction of five theses. The first, that "Psychological explanation is typically nomic and is intentional through and through" (p. 7), amounts to the denial of both eliminative materialism and anomalous monism. While neither of these doctrines is discussed in this work, the latter is considered at length (and

ultimately dismissed) in Fodor and Lepore (1992). The second thesis asserts that " 'Mental representations' are the primitive bearers of intentional content" (p. 7). Together with the third thesis, that "Thinking is computation" (p. 9), this claim entails the psychological functionalism familiar to readers of Fodor (1975, 1983). Somewhat more surprising is the inclusion of the fourth thesis, that "Meaning is information" (p. 12). "Information" is understood in the sense of the informational semantics of Dretske (1981); in Fodor's words, "what bestows content on mental representations is something about their causal-cum-nomological relation to the things that fall under them" (p. 12). The concept CAT means what it does in virtue of the (unspecified) causal-cum-nomological relation between its tokens (thoughts about cats) and cats. Fodor's endorsement of information semantics is well-known, and indeed a cornerstone of his *Theory of content* (1990); what is surprising here is not the presence of this fourth thesis *per se* but rather its inclusion *as a constituent of RTM*. Finally, the fifth thesis of Chapter 1 urges that "Whatever distinguishes coextensive concepts is ipso facto 'in the head' " (p. 15). By means of this thesis Fodor distances himself from temptations to import the Fregean program wholesale into information semantics.

Having set out a canonical version of RTM, in Chapter 2 Fodor proceeds to deduce from it five "non-negotiable" conditions that a theory of concepts consistent with RTM must meet. The first, that "Concepts are mental particulars" (p. 23), is a tenet of any *representational* theory, though not a claim acceptable to advocates of some alternative approaches to the mind, such as behaviorism. Fodor's second condition, that "Concepts are categories and are routinely employed as such" (p. 24), asserts a relation between concepts and the things which fall under them; "applications of concepts are susceptible of 'semantic evaluation' " (p. 24). The third, that "Concepts are constituents of thoughts and, in indefinitely many cases, of one another" (p. 25), requires that successful theories of concepts explain the systematicity and productivity of thinking, believing, and other propositional attitudes. Running the risk of baffling the reader familiar with his strong nativist tendencies, Fodor further insists that "Quite a lot of concepts must turn out to be learned" (p. 27). The third and fourth conditions together express the importance Fodor places on explaining the *compositionality* of concepts, a task at which, Chapter 5 will argue, prototype-based theories fail dismally. Finally, "Concepts are *public*; they're the sorts of things that lots of people can, and do, *share*" (p. 28). RTM, Fodor argues, is inconsistent with conceptual holisms which assert that no two subjects can share any belief unless they share all (or a great many) of their beliefs. Such views are the target of much more protracted attack in Fodor and Lepore (1992). Chapter 2 concludes with a reflection on the compatibility of the goal of saving the architecture of a Fregean theory of meaning with that of completing RTM. When Frege is stripped of his Platonism, psychologically constituted modes of presentation (hereafter MOPs) must take the place of Fregean senses in distinguishing co-extensive expressions. The search for such MOPs is thus the search for mental particulars of precisely the sort needed to complete RTM.

With these preliminaries out of the way, the stage has been set for an exposition and defense of informational atomism. Remaining true to the constraints of the

"what else" argument, Fodor must first refute the view he has set up as his chief rival. Interestingly enough, for Fodor there is at bottom only one such view, inferential role semantics (IR-semantics), though there are several variations. The two chief variations are that which treats concepts as definitions, and that which treats them as stereotypes or prototypes. For the last 20 years, the debate over the nature of concepts has behaved as though these positions were opposing poles. But for Fodor's purposes, the fact that both allow conceptual content to be determined, in part, by inferential role requires they be classed together. Chapters 3 and 4 of *Concepts* argue concepts can't be definitions, and Chapter 5 argues they can't be prototypes or stereotypes. Readers familiar with the history of both argument constellations will doubtless trace the former to Fodor *et al.* (1980) and the latter to Fodor's 1981 essay, "The present status of the innateness controversy."

Chapter 3 is devoted to criticizing recent efforts by linguists to save definitions. Jackendoff's (1992) attempt to account for polysemy (such as the apparent polysemy of "keep") is decried as circular when taken as a defense of definitions, since "keep" appears polysemic only when we assume it has a definition in the first place. What makes "keep" univocal, for Fodor, is not that all of its divergent tokenings somehow respect a common definition, but rather that they all refer to instances of *keeping*. This sort of disquotational move characterizes most of Fodor's reluctant forays into semantics. The target of the remainder of Chapter 3 is the developmentally motivated arguments of Pinker (1984, 1989). The "semantic bootstrapping" argument of 1984 is shown to establish, at best, the existence of lexical semantic features, but "an argument for lexical semantic features is not ipso facto an argument that there is lexical semantic decomposition" (p. 63). Since these arguments are tangential to our critical interests, we pass over them without further comment.

The title of Chapter 4, "The demise of definitions, Part 2: the philosopher's tale," is somewhat misleading. The familiar tale told by contemporary philosophers begins with Quine's attack on the analytic/synthetic distinction. This is not Fodor's tale, though it serves as background material. What he offers instead is a kind of psychopathology of the once prevalent philosophical conviction that some necessary connections among concepts were constitutive of those concepts, and hence yielded necessary conditions for concept possession. This conviction, we recall, lay at the heart of the philosophical belief in analytic truths about concepts. The assumption that concepts were constituted by their definitions, in turn, provided a nice explanation for the existence of such truths. Since Fodor denies the existence of analytic truths, the fact that a definitional structure for concepts would explain such truths fails to provide any evidence in support of definitions. However, our strong intuition that some necessary connections among concepts are constitutive of those concepts warrants at least some *psychological* explanation. Chapter 4 offers one.

Says Fodor, if you are in possession of the concept DOG "It's *that* your mental structures contrive to resonate [2] to *dog*hood, not *how* your mental structures resonate to *dog*hood, that is constitutive of concept possession according to the informational view" (p. 76). The *how* of this "resonance" is a matter of mere "semantic access." Some concepts permit many divergent avenues of semantic access. An instance of the concept VENUS, for example, may be detected by

observing either the first star to appear at dusk, or the last to vanish at dawn. Semantic access to other concepts is more restricted. If informational semantics is assumed, then with the notion of semantic access in hand, the illusion of analyticity may be explained as follows: consider concepts for which, like the one-criterion concepts of Putnam (1983), all avenues of semantic access diverge from a single trunk. One such concept is BACHELOR; determining that *S* is a bachelor always ultimately requires determining, one way or another, that *S* is an unmarried man.

> Suppose you think the only epistemic route from the concept C to the property that it expresses depends on drawing inferences that involve the concept C*. Then you will find it intuitively plausible that the relation between C and C* is conceptual; specifically, that you can't have C unless you also have C*. And the more you think that it is *counterfactual supporting* that the only epistemic route from C to the property it expresses depends on drawing inferences that involve the concept C*, the stronger your intuition that C and C* are conceptually connected will be (p. 83).

Our intuitions regarding constitutive conceptual relations, then, are artifacts of our epistemic habits. So, then, are our intuitions regarding analyticity. And if our intuitions regarding analyticity are mere artifacts, they provide no evidence for the thesis that concepts are definitions.

It is worth pausing at this point to note a source of potential confusion. The avenues of "semantic access" to concepts discussed in Chapter 4 can appear suspiciously akin to the MOPs (de-Platonized, re-psychologized Fregean senses) of Chapter 1. Both notions are meant to cover whatever largely psychological mechanisms mediate between subjects and the (informationally construed) contents of their thoughts. Yet while in Chapter 4 it is clear that a given concept may have seemingly arbitrarily many routes of semantic access, MOPs were supposed to play a role in *concept individuation* ("... MOPs can individuate concepts and referents can't ..." p. 19). Specifically, they were meant to distinguish between synonymous concepts, where for Fodor, concepts are synonymous if they bear the same information. Fodor appears to want it both ways. On the one hand, he wants the fine-grained individuation for concepts provided by MOPs, though of course the *contents* of these finely individuated concepts are still given entirely by the nomological relations between their tokenings and their instances. On the other hand, he wants the coarser-grained individuation provided by purely informational criteria. We will return to this apparent ambiguity and a problem that it yields later.

Chapter 5 canvasses a further variety of IR-semantics, that which treats concepts as stereotypes or prototypes. The crux of Fodor's argument against this proposal remains as in Fodor (1981): compositionality is a non-negotiable condition on theories of concepts, and prototypes don't compose. Fodor acknowledges the ample evidence for the psychological reality of prototypes, and even asserts, "The discovery of the massive presence of prototypicality effects in all sorts of mental processes is one of the success stories of cognitive science" (p. 93). So, (many) concepts have associated prototypes. But the attempt to identify concepts

with their associated prototypes is a non-starter so long as prototypes fail to meet the compositionality constraint.

The argument of this chapter divides into two parts: a defense of the compositionality constraint, and a demonstration that prototype accounts violate it. In the first part, Fodor somewhat impatiently reviews familiar reasons why concepts should be taken as compositional ("One does wonder, sometimes, whether cognitive science is worth the bother," p. 98). Of interest to those who have followed Fodor's career is the admission that the apparent productivity and systematicity of concepts fail to clinch the case for compositionality, though of course the latter would *explain* the former. The argument of Fodor (1975) is thus deemed inconclusive. The "best" argument for compositionality remains that its effects are in evidence in countless details of our cognitive capacity, as, for example, in our ability to refer to individuals by means of definite descriptions.

The argument that prototypes can't compose boils down to the observation that prototypes for complex concepts aren't inherited from the prototypes of their constituent concepts. Examples of this failure include "Boolean" concepts, such as negations, and the now classic PET FISH. Since Fodor's argument is doubtless familiar to many readers, and since we have no desire to take issue with it here, we here refrain from further discussion. Two additional dialects of IR-semantics, "meaning postulate" and "theory theory" approaches, are consigned to appendices to Chapter 5.

The task of discrediting various versions of IR-semantics having been accomplished, in Chapters 6 and 7 Fodor sets out his exposition and defense of informational atomism (IA). The exposition portion of the exposition and defense, however, is curiously short; in fact, it occupies all of four sentences at the beginning of Chapter 6:

> IA has an informational part and it has an atomistic part. To wit:
>
> —Informational semantics: content is constituted by some sort of nomic, mind–world relation. Correspondingly, having a concept (concept possession) is constituted, at least in part, by *being* in some sort of nomic, mind-world relation.
>
> —Conceptual atomism: most lexical concepts have no internal structure (p. 121).

The remainder of the book is devoted to addressing three objections, one of which (that IA makes conceptual analysis impossible) is simply dismissed. The remaining two claim that IA has absurd consequences: (1) that there are laws governing nearly *all* of our lexical concepts, including DOORKNOB, and (2) that nearly all of our lexical concepts, including DOORKNOB, are innate. Chapter 6 considers the latter, Chapter 7 the former.

On the issue of radical nativism, one might expect Fodor to simply bite the bullet. This, of course, was his response to the apparent nativist consequences of RTM in 1975 and 1981. And in *Concepts*, it sometimes looks as though Fodor *thinks* he has bitten the nativist bullet again; witness the opening of Chapter 7:

> Here's how we set things up in Chapter 6: Suppose that radical conceptual atomism is inevitable and that, atomism being once assumed, radical conceptual nativism is inevitable, too. On what, if any, ontological story would conceptual nativism be tolerable? (p. 146)

However, the discussion of Chapter 6 is considerably more complicated and subtle than this brief description might lead us to believe. Indeed, it closes with the remark, "Maybe there aren't any innate *ideas* after all" (p. 143). The details of this discussion will prove of considerable interest.

The inference from atomism to the innateness of all primitive (undefined) concepts, including DOORKNOB, is underwritten by what Fodor calls the Standard Argument (SA). The SA turns out to bear a notable resemblance to the argument of Fodor (1975), condensed into a single paragraph. Starting from the assumption that concept learning is an inductive process of hypothesis testing and confirmation, the SA proceeds by noting that primitive concepts can't be learned this way, on pain of circularity ("... to learn RED inductively you'd have to devise and confirm the hypothesis that things fall under RED *in virtue of being red*. But you couldn't devise or confirm that hypothesis unless you already had the concept RED ..." p. 124). It follows that primitive concepts must be unlearned, hence innate.

Rather than standing pat with this conclusion, Fodor notes, first, that the SA's account of concept learning assumes a cognitivist account of concept possession, and second, that IA entails a *non-cognitivist* account of concept possession. This observation permits him to attempt to strengthen his hand by divorcing IA from radical concept nativism; "... if you're prepared to settle for a theory of concepts that is plausibly *compatible* with the denial of radical nativism, maybe we can do some business" (p. 126).

To possess a given concept, for Fodor, is to be "nomologically locked" to the property expressed by the concept. It follows that to acquire a concept is to become thus locked. One might be tempted to conclude that IA blocks SA, since on the surface, there seems no reason to believe that becoming nomologically locked necessarily requires hypothesis testing. Fodor quickly disavows this move, however. The concept DOORKNOB presumably expresses the property *doorknob*. On the informational view, the fact that the former expresses the latter is explained by the assumption that, in one who already possesses the DOORKNOB concept, doorknobs cause DOORKNOB tokenings. DOORKNOB would not express *doorknob* if its tokenings were generally caused by something other than doorknobs. To acquire the concept is to enter into this non-arbitrary causal relation. The acquisition of DOORKNOB thus seems possible only given a *particular kind* of sensitivity to the causal potency of doorknobs: doorknobs must be treated as evidence. We use *evidence* to confirm or refute hypotheses. It thus emerges that IA's account of concept acquisition, cognitivist or not, must invoke hypothesis testing, thus running headlong into the SA. Recourse to selectionist gambits and the "mental triggering" language of Fodor (1981) proves no help.

In the absence of any alternative to the assumption that "the relation between concepts and experiences is typically evidential" (p. 132), a contradiction ensues:

> It is perhaps tolerable that representational theories of mind should lead by plausible arguments to a quite radical nativism. But it is surely not tolerable that they should lead ... to contradiction. If [the doorknob/DOORKNOB relation] shows that primitive concepts *must* be learned inductively, and SA shows that they *can't be* learned inductively, then the conclusion has to be that there aren't any primitive concepts. But if there aren't any primitive concepts ..., RTM has gone West. (pp. 131–132)

The goals of saving RTM and of demonstrating the compatibility of IA with the denial of nativism thus turn on the discovery of an alternative to the evidential construal of the relation between acquired concepts and experience. Fodor's strategy is to view the doorknob/DOORKNOB relation as "the consequence of a *metaphysical* truth about how concepts are constituted, rather than an empirical truth about how concepts are acquired" (p. 133). Two such approaches present themselves. The first, which treats the doorknob/DOORKNOB locking as a Kripke/Putnam style causal or historical relation, is dismissed almost immediately, on the grounds that not just any causal interaction with doorknobs will give one the DOORKNOB concept. The second proposal receives Fodor's endorsement: "Maybe what it is to be a doorknob isn't *evidenced* by the kind of experience that leads to acquiring DOORKNOB; maybe what it is to be a doorknob is *constituted* by the kind of experience that leads to acquiring the concept DOORKNOB" (p. 134). To be a doorknob is to be the kind of thing we (typically) experience as a doorknob, or in Fodor's words, "*being a doorknob* is having that property that minds like ours come to resonate to in consequence of relevant experience with stereotypic doorknobs" (p. 137). This account has the advantage of explaining not only the doorknob/DOORKNOB relation, but also the importance of stereotypicality effects in concept acquisition. It may entail a kind of nativism, but this nativism "is perhaps not one of *concepts* but of *mechanisms*" (p. 142), such as whatever mechanism underwrites sensitivity to stereotypical instances. We will return to consider some metaphysical consequences of this approach in our critical discussion.

This story having once been told, the reply to the second objection to IA ("How could their be *laws* about *doorknobs*?") becomes fairly predictable: the laws about doorknobs are in reality laws about *us*. This doctrine is expounded in Chapter 7. The idea that (most) concepts function as they do because they are constituted by locking to mind-dependent properties must be qualified in two important ways. First, says Fodor, it is important to recognize that, appearances to the contrary notwithstanding, this doctrine does *not* commit IA to idealism; since minds exist in the world, mind-dependent properties exist in the world. A further qualifier is required to allow IA to deal with natural kind concepts. For such concepts, it seems implausible to say that they are locked to mind-dependent properties. The difference between the folk concept WATER and its sophisticated theoretical counterpart is, at bottom, that while both are nomologically locked to the property *being water*, the lockings in question support different ranges of counterfactuals. For example, nomological locking of the folk concept WATER suffers a breakdown in worlds containing a substance which looks, feels, and tastes like water, but is in fact not

H_2O: the substance is (mistakenly) ruled an instance of WATER. Ultimately, for Fodor, there is only one concept WATER, since both the naif and the scientist are nomologically locked to the same property in their environment. But their locking is supported by different routes of semantic access (see above discussion of Chapter 4). In the case of a scientist, semantic access is mediated by a theory of water, and perhaps even by a metatheory in which such concepts as NATURAL KIND feature prominently. But as before, Fodor insists that content is determined by the *fact* of nomological locking, not the *how* of semantic access.

3. Critical discussion

We begin by canvassing several related, highly compelling objections to which Fodor seemingly has a ready reply. First, consider the issue of natural kind concepts, like WATER. Fodor claims (in Chapter 7) that the difference between the naif and the scientist is that, while both may have the concept WATER, they are nomologically locked to *water* to different degrees. Their locking supports different counterfactuals, and in particular, the naif is a poor water detector in those worlds (like Twin Earth) in which substances other than H_2O exhibit the superficial characteristics of water. Now, one might object, who is to say that the naif possesses a concept which locks him or her to H_2O in only some possible words, rather than one which locks him or her to the disjunction of H_2O and XYZ in a much larger range of possible worlds? Alternatively, how are we to tell the difference between a careless sophisticate and an earnest naif?

A similar constellation of objections raises the broader issue of representational error, hearkening back to the discussion of Fodor (1990). A subject *S* identifies a robotic sheep as an instance of SHEEP. How are we to know that *S* is locked (imperfectly) to *sheep*, and not to the disjunction of *sheep* and *robotic sheep*? More generally, how do we know that *S*'s concept X is locked to the property *x*?

Fodor's reply, hinted at throughout *Concepts* and in the earlier "Theory of content" essay, is that questions of the form "How do we know X?" are *epistemological* questions. RTM and its associated theory of concepts, however, are *metaphysical* doctrines. Metaphysics is independent of, and perhaps in some important sense prior to, epistemology. In response to the question, did Homer have "the same concept of water we do," Fodor asserts, "I don't much care which you say, so long as you like the general picture" (p. 157). The implication is that Fodor's theory of concepts entails that there *is* a fact of the matter as to which ambient property Homer was locked to, and that this fact persists *whether or not it is known* (or even knowable). Or consider the following passage from Fodor (1990):

> "What makes you so sure that the counterfactuals are the way that you're assuming? Who says that [a frog's snapping at flies is] asymmetrically dependent on [its snapping at black dots] and not vice versa?" Strictly speaking, this is a sort of question I do not feel obliged to answer; it suffices, for the present metaphysical purposes, that there are naturalistically specifiable conditions, not known to be false, such that *if* they obtain

> there is a matter of fact about what the frog is snapping at. (Fodor, 1990, p. 107).

For a metaphysical theory to entail that there are unknown facts is neither particularly surprising nor particularly disturbing. Somewhat more disturbing, from our perspective, is the claim that there is a determinate fact of the matter, even when that fact is *unknowable*. And it strikes us that if Fodor is right about concepts, our situation with regard to the questions asked above is more like the latter. To determine whether a given subject is nomologically locked to a given property would require, first, some information on the range of possible worlds in which the subject's tokenings of a particular concept are positively correlated with the property's presence, negatively correlated with its absence, and divorcable from the presence or absence of other properties. But further, it would require a criterion for distinguishing imperfect, frangible lockings, broken by the perversity of certain worlds, from mere correlations, unworthy to be called nomological lockings in the first place. None seems to be on offer.

However, Fodor is not insensitive to the presumption against theories that entail untestable facts. In his defense, it might be said that accepting these consequences of the theory of concepts is justified by the importance of the theory in completing RTM, the acceptance of which is justified, in turn, by its great explanatory power. To further press the issue of the proper order of dependencies in the relation between epistemology and metaphysics would go beyond the scope of this paper, though we note in passing that contemporary physical theories (*viz*. quantum mechanics) have turned away from positing unknowable facts. For the present, we propose to allow Fodor to beg off from such questions as, "How can we know that a subject *S* has concept X in virtue of being nomologically locked to *x*?"

That, however, is far from the end of the story when it comes to Fodor's bracketing of matters epistemological. Fodor has distanced himself not only from the general Theory of Knowledge, but also from a broad swath of the philosophy of psychology. It is this latter distancing we find most objectionable, and to which the remainder of this paper will be devoted. In what follows we will argue that whatever its merits as a theory of content *tout court*, IA is thoroughly unsuitable as a theory of *mental content*, hence as a theory of concepts conceived as *mental* particulars. The reason for this failing is that its metaphysics rests on a promissory note redeemable only with an answer to the question, "What is a mind?" More specifically, "How can metaphysical content be mental content?" or "How can a mind have access to the content contained in Fodor's informational atoms?" Fodor may wish to restrict himself to a consideration of metaphysical content, at least as a strategic move, but if his *metaphysics* of content makes the questions about *mental* content unanswerable, then the metaphysics itself is impeached. He may wish to *postpone* epistemological issues, but his metaphysics must not make them *necessarily inscrutable*. Further, the only kinds of answer to this question which have any hope of meeting our needs are those which turn on *how* minds do what they do—on explanations from the realm of functional psychology. But if such an approach to the mind is required to complete the theory of concepts, then informational accounts of conceptual content

look pretty hopeless. It turns out to be impossible, for example, for the informational atomist to provide a fully naturalized model of representational error. A (deeply) related problem is that some minds, at least some of the time, can fallibly detect their own representational errors—error with respect to content—and IA has no resources for explaining that possibility.

As advertised, much of what we have to say will apply equally well to the earlier Fodor (1990) and to work by Fodor's metaphysical allies. Following some preliminaries, we raise and consider two related questions. The first asks how fine-grained concepts are on Fodor's view. The second asks what it is for an organism or other entity to be in possession of mental content.

So as to avoid the risk of appearing to criticize Fodor for failing to meet goals his project was never meant to pursue, it is worth emphasizing the extent to which *Concepts* is intended as a theory of *mental* content. The very first of the five "non-negotiable conditions" Chapter 2 places on a theory of concepts is that it treat concepts as *mental particulars*. Further, we are told throughout that concepts are the constituents of *thoughts*. Finally, near the book's conclusion, Fodor reflects, "I've assumed throughout that informational semantics is, if not self-evidently the truth about *mental content*, at least not known to be out of the running" (p. 146). This point is worth stressing because it shows that one avenue of retreat open to defenders of other versions of informational semantics is not open to Fodor; it is not open to him to claim that he is interested only in some broader metaphysical notion of content, not mental content.

We are now in a position to raise our primary objections.

3.1. How fine-grained are concepts?

In our discussion of Chapter 4 (see above), we noted an apparent ambiguity in Fodor's treatment of the issue of concept individuation. Recall that, in Chapter 1, Fodor was concerned to salvage the Fregean explanation of the failure of substitutions of co-referential expressions in intensional contexts (see Frege's well-known Morning Star/Evening Star case). A given (informationally construed) content may have many modes of presentation (MOPs), the informational equivalence of which is often ignored by human subjects. Objecting to Frege's Platonism about senses, Fodor proposed that MOPs be taken as "in the head." Since informational semantics implies that co-referential concepts are synonymous, and since MOPs distinguish co-referential concepts if anything does, the fifth thesis of Fodor's first chapter, that "Whatever distinguishes coextensive concepts is ipso facto "in the head" follows straightforwardly.

In order to save Frege's solution to the problem of substitutivity failure while maintaining an informational theory of content, Fodor needs to allow, first, that two distinct concepts can be synonymous, and second, that mental entities, MOPs, are what distinguish synonymous concepts. In Chapters 1 and 2, Fodor thus gives every appearance of having a theory of concepts on which they are at least as fine-grained as the meanings of English expressions.

Chapters 4 and 7, however, present a much coarser-grained picture of concept

individuation. Fodor's explanation of the illusion of analyticity (and hence of the superficial plausibility of the definitional view of concepts) turns on there being multiple routes of semantic access to a given *concept*—not to a given *content* (MOP-individuated concepts provided multiple routes of semantic access to a given content)—recall it was one-criterion concepts that tricked us into thinking they had necessary constituents.

A problem arises, however, concerning the relationship between MOPs and routes of semantic access. Fodor's MOPs certainly sound like Frege's MOPs, and Fodor's routes of semantic access sound like Frege's routes to the referent, but, for Frege, both are aspects of sense, and, therefore, are either in some sense identical or at least in one-to-one correspondence. Fodor has disavowed Frege's Platonism, but he does not address this Fregean linkage between MOPs and routes of semantic access via senses, neither to disavow it nor to show that he is not committed to it. Unfortunately, there are serious consequences if this linkage is maintained—or cannot be avoided.

Now, the set of routes of semantic access and the set of MOPs need not be identical. In particular, it is possible to conceive of non-intentional, extra-mental routes of semantic access. But at a minimum, MOPs would seem to qualify as routes of semantic access were it not for their concept-constitutive role. From Fodor's perspective, a more promising way of distinguishing MOPs from routes of semantic access might seem to be to allow multiple routes of semantic access to a given MOP-individuated concept. But Fodor himself appears to rule out this line of attack in his efforts (in Chapter 1) to distance himself from the Fregean program. Since MOPs are destined to play a role individuating concepts *qua* mental particulars, he is especially concerned to reject the Fregean thesis that "MOPs are abstract objects; hence they are non-mental" (p. 16). Toward this end, he insists that "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*" (p. 17). This insistence is motivated by the following analogy (see p. 18): one might reasonably call a diagram (say, of a triangle) a mode of presentation of a property (say, *triangle*). But an MOP of this kind can't individuate concepts, for a given diagram might be used to present very different contents on different occasions. The same holds for Fregean senses, construed as abstract, non-mental objects [3]. So in order for MOPs to individuate concepts, there must be exactly one way of grasping (or entertaining) a given MOP.

The distinction between MOPs and routes of semantic access is thus *prima facie* ambiguous. It does not immediately follow, however, that this ambiguity poses any serious problem for Fodor's theory of concepts. On the surface, there appear to be two ways of resolving the ambiguity. Fodor might give up on MOPs as concept-individuating mental entities. Or he might abandon the claim that a given concept may have multiple routes of semantic access. Both horns turn out to have unacceptable consequences.

Suppose Fodor took the first horn, abandoning the thesis that mental MOPs distinguish synonymous concepts. He would, of course, have to set aside the project of saving Frege's explanation of substitutivity failure, a significant but

ultimately expendable goal of IA. The following observation would, of course, remain unaddressed:

> The Frege programme needs something that is both in the head and of the right kind to distinguish coreferential concepts, and the Mates cases suggest that whatever is able to distinguish coreferential concepts is apt for syntactic individuation. Put all this together and it does rather suggest that modes of presentation are syntactically structured mental particulars. (p. 38)

But perhaps more importantly, abandoning concept-constitutive MOPs would undermine our confidence in the close similarity of the structure of thoughts with the surface structure of English sentences, or to use Fodor's words, it would undermine our confidence that "individuating MOPs [and hence concepts] is more like individuating forms of words than it is like individuating meanings" (p. 17). And it is only such confidence that makes Fodor's apparent assumption that the answers to most semantic questions are disquotational remotely plausible (see Chapter 3, *passim*). We conclude that the first horn is unappealing.

To instead abandon the pairing of individual concepts with multiple routes of semantic access would, of course, involve discarding Fodor's explanation for the illusion of analyticity. This is a serious price to pay, but affordable when taken by itself, since there are plenty of reasons for doubting that concepts are definitions even in the absence of an explanation for why they *looked* to us so much like definitions for so long. More significantly, however, the second horn demands that we abandon Fodor's account of the difference between the naif and the scientist. In Chapter 7, it is important to Fodor to claim that both have the same concept WATER, in that both are nomologically locked to the property *being water*. The fact that the scientist's locking is preserved over a broader range of possible worlds is explained by recourse to semantic access: the scientist's semantic access to water is mediated by a theory. This explanation relies on the pairing of MOP-individuated concepts with multiple routes of semantic access [4].

In short, the ambiguity over the fineness of conceptual grain noted in our discussion of Chapter 4 turns out to have serious and far-reaching consequences. It poses a dilemma, either horn of which severely hedges some of the virtues touted for IA. We will return to discuss an aspect of this dilemma in our treatment of our second primary objection.

3.2. What is it for an organism (or other entity) to be in possession of mental content?

Our reading of *Concepts* leads us to this question by two different routes. First, we are intrigued by the conclusion of Chapter 6 that RTM commits us to a nativism of *mechanisms* rather than a nativism of ideas. This claim is interesting in and of itself, since it represents a departure from Fodor's long-standing commitment to what has been called (e.g. in Stich, 1968) *dispositional* nativism, and a move toward a *functional* nativism more akin to that of Noam Chomsky. But further, the suggestion that there are innate mechanisms necessary to acquiring or having concepts implies

that, in an important sense, what it is to be a mind, or to be in possession of mental content, is just to be possessed of these innate mechanisms in an environment suitable for their deployment. One version of our present question, then, is "What is the nature of the innate mechanisms necessary for concept acquisition and possession?"

A further approach is spurred by Fodor's remark that "having a concept is something like 'resonating to' the property that the concept expresses" (p. 137). Words to this effect, sometimes with "being a detector for" used in place of "resonating to," may also be found throughout Fodor (1990). This claim seems to us to seriously underdetermine concept possession, as witnessed by a counterexample we will develop shortly. Having at least one concept is clearly a minimal necessary condition for possessing mental content. To scrutinize what it means to have at least one concept is thus also to address the guiding question of this section.

We turn first to the innate mechanisms necessary for concept possession. In Chapter 6, Fodor makes a claim of considerable importance:

> ... all that needs to be innate for RED to be acquired is whatever the mechanisms are that determine that red things strike us as they do; which is to say that what needs to be innate is the sensorium ... What has to be innately given to get us locked to *doorknobhood* is whatever mechanisms are required for doorknobs to come to strike us as such ... the kind of nativism about DOORKNOB that an informational atomist has to put up with is perhaps not one of *concepts* but of *mechanisms*. (p. 142).

We recall that Chapter 6 had argued that the nomological locking of subjects possessing DOORKNOB to the property *doorknobhood* ought to be explained *metaphysically*, not psychologically; the property *doorknobhood* just is the property of striking us in a certain way. DOORKNOB is to be understood by analogy to RED, a concept which locks us to a property constituted by the make-up of the human sensorium. This analogy sets up the sensorium as the paradigm case of an innate mechanism necessary for concept possession.

As mentioned in our expository discussion of Chapter 6, we think the shift from a nativism of concepts to a nativism of mechanisms represents an important departure from Fodor's earlier views. It sometimes appears as though Fodor thinks so, too. At any rate, one consequence of this view is worth noting by way of a segue into our detailed discussion of the question of how IA accounts for mental content. For the informational atomist Fodor of the late 1990s turns out to face a very different sort of problem from the self-styled "mad-dog nativist" Fodor of the 1970s and 1980s. The main problem for the latter was always the *prima facie* implausibility of the view that the vast majority of concepts, and in all likelihood DOORKNOB, too, had to be innate. But implausible as this account may have seemed, it at least offered the rudiments of a story about mental content. The mental contents of our concepts were in the head, where they belonged, and there was no mystery as to how they got there (they were innate), though of course considerable mystery as to the conditions under which one or another innate concept came to be expressed.

But while the nativism of mechanisms espoused in *Concepts* may represent an

improvement over the old view in terms of *prima facie* plausibility, any such improvement is bought at the price of a coherent account of mental content. All that Fodor has to say about the mental content of a subject *S*'s concept DOORKNOB is that the sensorium of *S* is so constituted as to allow *S* to come to resonate with doorknobs, making *S* a (potentially) reliable doorknob detector. We now turn to the inadequacy of this proposal.

Mental content? What makes something a mind is not only what it does but how it does it. On any version of RTM, to be a mind is to think, and to think is to manipulate concepts in the sense of mental particulars. To manipulate concepts one must *have* concepts. Fodor claims that "having a concept is something like 'resonating to' the property that the concept expresses" (p. 137). The obvious counterexamples would appear to be "dumb" detectors, devices which reliably "resonate to" certain properties but which fail to exhibit any mentality. In Chapter 6, Fodor considers one sort of dumb detector (hereafter DD), an engineering construct designed to sort the doorknobs from the non-doorknobs. In this case, the DD is nomologically locked to a property, but its locking is derived in the sense that it depends on the engineers' being nomologically locked to doorknobs. Fodor is thus correct that it fails to provide a counterexample.

But of course there are naturally occurring DDs. Consider the following case. The dendrites of a nerve cell contain receptors to which molecules of a specific neurotransmitter, say dopamine, may bind. When enough dopamine molecules have bound to the cell's receptors, the cell fires; otherwise, under most circumstances, it remains quiescent (but see below). Clearly, the cell is, or has, a dopamine detector. It is nomologically locked to the ambient property *dopamine-hood*, to which it reliably resonates.

In this DD case, our intuitions suggest to us that while information may be present, there is no *mental* content; the cell does *not* have the concept DOPAMINE. Nor does it do any good to argue that the cell must meet the minimal conditions for *functional* content, that the information contained in its own internal states (on the presence or absence of dopamine) must be used in further processing. The cell meets this condition, too, because the state of its dopamine detector *is* functional for the cell, yet it still lacks mental content.

For that matter, it still lacks mental content if we flesh out its description to include a kind of story about error. Suppose that the cell's environment can be made to contain a kind of poison, call it "crank," one which binds to its dopamine receptors, causing the cell to fire even in the absence of dopamine. Had we been tempted to say, in the first place, that the cell had the concept DOPAMINE, we could now talk about its representational errors. We could say that the cell has the concept DOPAMINE, and not the concept DOPAMINE OR CRANK, because its functioning as a crank detector is *asymmetrically dependent* on its functioning as a dopamine detector (see Fodor, 1990). The existence of the nomological relation between dopamine and DOPAMINE tokens, one might say, is a necessary condition for the existence of the nomological relation between crank and DOPAMINE tokens.

This and similar DDs meet all of Fodor's conditions: they contain information in the form of nomological lockings to properties in the environment, and those lockings respect asymmetric dependence. Yet they lack both concepts and mental content (Bickhard, 1993, 1996). Further, it seems difficult to see how Fodor *could* rule DDs out of court, even by adding additional conditions. As he points out in Chapter 4, "There is nothing in informational semantics that stops content-making laws from being basic. For that matter, I suppose there's nothing in metaphysics that stops *any* law from being basic ..." If Fodor is right in this, then IA is inadequate as a theory of *mental* content. It also follows that the formulation of such a theory is not a task for metaphysics alone, but also for psychology; to rule out DDs as having mental content, we must allow the nomological lockings constitutive of concept possession to proceed by way of *psychological* laws.

Representational error. To disallow the DD cases as instances of mental content thus requires a theory of what it is to be a mind that incorporates an account of *how* minds, *qua* minds, detect the properties to which they lock. Among other mental properties regarding representations, minds, like DDs, have a capacity to err, but unlike DDs, at least some minds can sometimes detect their own representational errors. That is, mental representations, at least *some* of them, for some animals, are not only capable of *being* in error, such a condition of error is also *detectable*, at least fallibly, by those same animals—the class of which obviously includes human beings. Fodor's asymmetric dependency conditions not only fail to adequately rule out DDs, but they are also determinable in a particular case, if at all, only by rather sophisticated observers of the organisms and their environments. In particular, they are not determinable by the organisms themselves; therefore, error, as thus modeled, is not determinable by those organisms. Without organism- or system-detectable error, neither error-guided behavior nor error-guided learning is possible, yet both clearly occur. Fodor's metaphysics for content, then, cannot account for system-detectable error. Fodor's model has a long way to go to be able to account for *mental* content.

Naturalized error. Fodor could again rule such considerations irrelevant because they are not metaphysical—though that would leave his metaphysics for representation inadequate for *mental* content, and, therefore, of questionable *relevance* as a metaphysics. But there is an additional take on the importance of system-detectable error that applies even more directly to Fodor's metaphysics. Fodor has recognized the relevance of some potentially non-metaphysical considerations to his metaphysics. His attention to the asymmetric dependence device (Fodor, 1990) for example, reveals his desire to show how representational error is possible at all given an informational semantics. By criteria he himself endorses, if Fodor's approach makes any naturalized account of system detectable error *impossible*, his metaphysics is thereby impeached. After all, what we want, what Fodor wants, is to "[link] a theory of meaning with a theory of mind" (Fodor, 1987, p. 81). We need a theory, or at least Fodor needs a theory, not only of what it means for something to *contain* information, but also of the *encoding* of that information for a mind, of making that

information available. But "not every situation encodes the information that it contains" and "we haven't got a ghost of a Naturalistic theory about [encoding]" (Fodor, 1987, pp. 80–81). "What we're all doing is really a kind of logical syntax (only psychologized); and we all very much hope that when we've got a reasonable internal language (a formalism for writing down canonical representations), someone very nice and very clever will turn up and show us how to interpret it; how to provide it with a semantics" (Fodor, 1981, p. 223).

As Fodor's remarks suggest, the requirement that a metaphysics of content account for the possibility of error interacts with the requirement of naturalism. Naturalizing representation requires naturalizing representational normativity, which, in turn, requires naturalizing representational error. Fodor's account of the possibility of representational error in terms of asymmetric dependencies is naturalistic in the sense that it is rendered in terms of factual and counterfactual relationships, but it is relatively silent with respect to normativity, thus raising once more the question of whether Fodor has succeeded in accounting for the possibility of error at all.

The standard approach for accounting for normativity is in terms of various evolutionary historical considerations. Fodor might be willing to account for the *origins* of content-forming mechanisms in the sensorium in terms of some sort of evolutionary history—so long as the account does not commit the sin of adaptationism—but he does not want to account for the *metaphysics* of that content in terms of such a history. The problem is that he skips over the normativity issue, and that is the issue that pulls most strongly in the evolutionary direction in which Fodor does not want to go. We do *not*, in fact, endorse the standard evolutionary move for modeling function and functional normativity (Bickhard, 1993), but the point here is that Fodor has not adequately attended to this essential aspect of his metaphysics, one that threatens to undermine his entire project [5]. The naturalization of representational normativity thus emerges as a dangerous metaphysical gap in Fodor's theory.

We are suggesting that we *still* do not have a ghost of a naturalistic theory of encoding, of available representational content. Furthermore, the thrust of the points about a naturalized normativity of representational error and of system-detectable representational error is that Fodor has not addressed considerations which lead us to think that such accounts are *not possible* within the metaphysical framework that he develops (Bickhard, 1993). His metaphysical project might not be undermined by its current failure to provide such accounts, but it is refuted if it makes such accounts impossible.

System-detectable representational error. The issue of system-detectable representational error, then, cuts two related ways: (1) some minds do detect such errors at least some of the time—error guided behavior and learning require it—so any model that renders such detection impossible is thereby refuted as a model adequate for *mental* content; and (2) a purported metaphysics for representation that accepts the criterion of naturalism must naturalize representational normativity, and must therefore naturalize representational error. A model that can account for error only from the perspective of an external observer has not naturalized error: even if it were to get

the extension right—and that, of course, is itself contentious for Fodor's model (Loewer & Rey, 1991)—it fails to capture the normativity of error.

Fodor rejects issues of epistemology as being relevant to the metaphysical project in which he sees himself engaged. This rejection is explicit with respect to the sense of epistemology that refers to the epistemology of observers wanting to determine facts about humans' and other animals' concepts and contents. It is not as clear that Fodor wishes to reject the relevance of the epistemological conditions on the knowledge humans and animals possess concerning the contents of their own minds. On the one hand, the issues of how mental concepts and content are naturalistically realized might be postponed—arguably, necessarily postponed—until the metaphysics of what is *being* realized are clear. On the other hand, any purported metaphysics for concepts and content that renders that naturalization *impossible* thereby fails an even more fundamental metaphysical criterion.

Notes

[1] Fodor offers an argument for this privileging of metaphysics over epistemology. It seems that "people who start with 'what is concept *possession*?' [instead of "what are concepts?"] generally have some sort of Pragmatism in mind as the answer" (p. 3). Pragmatism, of course, leads to all sorts of terrible things, such as dispositional reductive analyses, or even behaviorism or verificationism. We might agree with Fodor about the advisability of privileging metaphysics if these slides from epistemological considerations to, for example, verificationism via pragmatism, were really logically forced, but they seem to be products of (historically common) misinterpretation rather than products of conceptual necessity.

[2] Fodor's adoption of the Gibsonian metaphor of resonance in the published version of his text (though not in the lectures on which the book is based) is curious. We wonder what further significance this choice has, if any, in the evolution of Fodor's thought.

[3] Fodor might claim that his route of semantic access and Frege's route to the referent are not the same things: they may involve the same path, but they have opposite directions of perspective, one from the sense to the referent, the other from the referent to the mental particular. Among other consequences, this move would entail that Fodor's metaphysical contents are *not* mental contents.

[4] Fodor's treatment of the difference between the naif and the scientist involves comparing concepts, and thus MOP-individuated concepts, *intersubjectively*, and that raises its own interesting questions. Either MOPs are shared, in which case we encounter the problems discussed above, or they are not, in which case *concepts* are not shared. And by the way, how could MOPs *possibly* be shared?

[5] There are multifarious additional problems with the general approach that Fodor advocates (Bickhard, 1993, 1996; Bickhard & Terveen, 1995).

References

BICKHARD, M.H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence*, *5*, 285–333.

BICKHARD, M.H. (1996). Troubles with Computationalism. In W. O'DONOHUE & R.F. KITCHENER (Eds) *The philosophy of psychology*, (pp. 173–183). London: Sage.

BICKHARD, M.H. & TERVEEN, L. (1995). *Foundational issues in artificial intelligence and cognitive science—impasse and solution*. Amsterdam: Elsevier Scientific.

DRETSKE, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

FODOR, J.A. (1975). *The language of thought*. New York: Crowell.

FODOR, J.A. (1981). *RePresentations*. Cambridge, MA: MIT Press.

FODOR, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
FODOR, J.A. (1984). Observation reconsidered. *Philosophy of Science*, *51*, 23–43.
FODOR, J.A. (1987). A situated grandmother? *Mind and Language*, *2*, 64–81.
FODOR, J.A. (1988). A reply to Churchland's "Perceptual plasticity and theroretical neutrality". *Philosophy of Science*, *55*, 188–198.
FODOR, J.A. (1990). *A theory of content*. Cambridge, MA: MIT Press.
FODOR, J.A., GARRETT, WALKER & PARKES (1980). Against definitions. *Cognition*, *8*, 1–105.
FODOR, J.A. & LEPORE, E. (1992). *Holism: a shopper's guide*. Cambridge: Blackwell.
JACKENDOFF, R. (1992). *Languages of the mind: essays on mental representation*. Cambridge, MA: MIT Press.
LOEWER, B. & REY, G. (1991). *Meaning in mind: Fodor and his critics*. Oxford: Blackwell.
PINKER, S. (1984) *Language learnability and language development*. Cambridge, MA: Harvard University Press.
PINKER, S. (1989). Learnability and cognition: the acquisition of argument structure, Cambridge, MA: MIT Press.
PUTNAM, H. (1983). "Two dogmas" revisited. In *Philosophical papers III: realism and reason*, Putnam's Collected papers (87–97). Cambridge: Cambridge University Press.
STICH, S. (1968). *Innate ideas*. Berkeley: University of California Press.