2

Has Content Been Naturalized?

LYNNE RUDDER BAKER

The Representational Theory of the Mind (RTM) has been forcefully and subtly
developed by Jerry A. Fodor. According to the RTM, psychological states that
explain behavior involve tokenings of mental representations. Since the RTM is
distinguished from other approaches by its appeal to the meaning or "content" of men-
tal representations, a question immediately arises: by virtue of what does a mental
representation express or represent an environmental property like *cow* or *shoe*?

This question asks for a general account of the semantics of mental
representation. Fodor places two conditions on the requisite theory: it must be
*physicalistic* (that is, it must be couched in nonsemantic and nonintentional terms,
free of expressions like "refers to" or "denotes" or "means that"), and it must be
*atomistic* (that is, it must allow that the thinker can have a single intentional state
without having any others). What is wanted, then, is a reductive theory that
"naturalizes" content by specifying sufficient conditions, in physicalistic and
atomistic terms, for a mental symbol to represent or express a certain property.

The Reduction

The naturalistic relation that carries the weight of the reduction of representation
is causation. The aim is to show how representational properties (e.g., the
property of representing *cat*) can be understood in terms of wide causal properties
(e.g., the property of being caused by *cat* instantiations, or, for short, by cats).
Fodor understands representation in terms of nomic relations between instances
of the property (cats, say) and mental tokens of a given type.

The basic idea is that tokens of a certain type represent those properties whose
instantiations produce them. The difficulty with this simple formulations is that
tokens of any given type may have countless different kinds of causes. Fodor calls
this phenomenon the "robustness" of thought. A thought of a cat may be
produced, for instance, not only by an instantiation of a cat, but also by an
instantiation of a shoe that you mistake for a cat, or by some preceding thought
of tigers, say, when there are not any cats around. Yet, we do not want to say that
your token represents some disjunctive property, *cat-or-shoe-or-preceding thought-
or-...*, and so on. We need sufficient conditions that allow your thought to
represent a cat and that rule out the disjunctive property.

To avoid this problem, which Fodor sometimes calls “the disjunction problem,” he formulates a notion of asymmetric dependence. Roughly, no matter what the actual cause of your token, it represents the property cat if non-cat-caused tokens of that type are “asymmetrically dependent” on cat-caused tokens of that type — if, that is, noncats would not cause tokens of that type unless cats did, but cats would still cause tokens of that type even if noncats did not. Asymmetric dependence, officially, is this:

(AD) The law C ‘\(\rightarrow\) D’ is asymmetrically dependent on the law ‘A \(\rightarrow\) B’ (where A, B, C, D are properties) iff the A/B connection can not be broken without breaking the C/D connection; but the C/D connection can be broken without breaking the A/B connection.

The asymmetric dependence of ‘C \(\rightarrow\) D’ on ‘A \(\rightarrow\) B’ is to be determined by answering the question — In the nearest possible world in which the A/B connection is broken, is the C/D connection thereby broken? — in the affirmative, and by answering the question — In the nearest possible world in which the C/D connection is broken, is the A/B connection thereby broken? — in the negative.

Combining the idea of asymmetric dependence with that of nomic relations, we have Fodor’s reduction, (R):

(R) A token of some nonsemantic type T represents a property P if
(i) instances of P cause (or are nomically related to) tokens of T, and
(ii) any tokens of T that are caused by instances of non-P are asymmetrically dependent on tokens of T that are caused by instances of P.\(^3\)

Clause (i) of (R) marks this view as a variety of information-based semantics, and Clause (ii) of (R) aims to solve the disjunction problem. As Fodor illustrates, “‘Cow’ means cow and not cat, or cat-or-cow, because there being cat-caused ‘cow’-tokens depends on there being cow-caused ‘cow’-tokens, but not the other way around.” Or, again, “[W]hat the story about asymmetric dependence comes down to is that ‘cow’ means cow because (i) there is a nomic relation between the property of being a cow and the property of being a cause of ‘cow’-tokens; and (ii) if there are nomic relations between other properties and the property of being a cause of ‘cow’-tokens, then the latter nomic relations depend upon the former” (TC, 40). (R), I think, captures these conditions.

I shall discuss two major tests of this account of representation: first, how it handles representation of uninstantiated properties, and second, how it solves the disjunction problem. Then, I shall raise a general theoretical problem for the reduction — a problem of specifying which relations are semantically relevant. Before turning to these issues, however, we must look again at (R). For I believe that Fodor vacillates in his interpretation of (R).

Interpretations of (R)

The first clause of (R) is open to different interpretations: Does it require that instances of P actually cause T-tokens, or that instances of P would cause T-tokens but for accidental circumstances (i.e., P is locally instantiated, and it does not matter whether or not P has caused T-tokens), or only that instances of P would...
cause T-tokens if the property P were instantiated (i.e., it does not matter whether P is locally instantiated or not). These distinctions lead to three different versions of (R). Call them the “actual-history” version, the “local-instantiation” version, and the “pure-informational” version, respectively.6

(AH) A token of nonsemantic type T means X if:
(1) “Xs cause T-tokens” is a law;
(2) For all Y (not = X), if “Ys cause T-tokens” is a law, then Ys cause T-tokens is asymmetrically dependent on Xs cause T-tokens;
(3) Some T-tokens are actually caused by Xs.

(LI) A token of nonsemantic type T means X if:
(1) “Xs cause T-tokens” is a law;
(2) For all Y (not = X), if “Ys cause T-tokens” is a law, then Ys cause T-tokens is asymmetrically dependent on Xs cause T-tokens;
(3) The property X is locally instantiated.

(PI) A token of nonsemantic type T means X if:
(1) “Xs cause T-tokens” is a law;
(2) For all Y (not = X), if “Ys cause T-tokens” is a law, then Ys cause T-tokens is asymmetrically dependent on Xs cause T-tokens.

Fodor’s official view is the pure-informational version. This line is in accord with his sustained emphasis on nomic relations between properties (which hold whether the properties are instantiated or not), as opposed to actual causal interactions among individuals. However, as we shall see, it is the local-instantiation version to which Fodor actually appeals (except when he is considering unicorns). Although he formulates the actual-history version as a possible alternative,7 he does not actually endorse it. And he does not distinguish the local-instantiation version from the others at all. Now, turn to the test cases.

Uninstantiated Properties: The Unicorn Case

We can represent uninstantiated properties like unicorn; and assuming it to be nomically possible that there be unicorns, we may represent unicorns by means of primitive symbols. (If you disagree, select any other uninstantiated property that might have been instantiated.) The asymmetric dependence view should allow that certain tokens represent unicorns.

Fodor has explicitly applied his view to the unicorn case. Clearly, neither the actual-history version nor the local-instantiation version permits a primitive symbol of Mentalese to mean unicorn. So, if Fodor’s view is to allow primitive symbols to represent uninstantiated properties, the view (as Fodor prefers) must be given the pure-informational interpretation.

Does the pure-informational version allow a person S’s U-tokens (internal tokens of some nonsemantic type U) to represent unicorn? The first clause of the reduction sails through fine: there is a nomic relation between unicorns and S’s U-tokens, because, as Fodor says, if S were in a world in which there are unicorns, they would cause S’s U-tokens.
But now consider: there is as much a nomic relation between S’s U-tokens and shunicorns – where a shunicorn is a unicorn look-alike that is really a small zebra with a horn in the middle of the forehead – as there is between S’s U-tokens and unicorns. The description of shunicorns is merely a heuristic device; ‘shunicorn’ is as primitive a term as ‘unicorn.’ The fact that ‘unicorn’ is an English word, but ‘shunicorn’ is not, is strictly irrelevant to the case. If either unicorns or shunicorns could have been instantiated in our world, then both could have been.

Since shunicorns are instances of non-unicorns, the asymmetric dependence condition comes into play. The asymmetric dependence condition should allow misrepresentation of a shunicorn as a unicorn, but it does not: Since we have no basis on which to distinguish the relative distances of worlds in which there are shunicorns and worlds in which there are unicorns from the actual world, we should treat them as equidistant from our world. Similarly, worlds in which a person’s shunicorn/U connection is broken, but her unicorn/U connection remains intact are the same distance from the actual world as worlds in which her unicorn/U connection is broken, but her shunicorn/U connection remains intact. At least, we have no principled way to distinguish the two.

The fact that worlds in which unicorns cause U-tokens, but shunicorns don’t, seem equidistant from us as worlds in which shunicorns cause U-tokens, but unicorns don’t, suggests a general dilemma concerning un instantiated properties: either the required asymmetric dependence is missing and the pure-informational version (along with the other versions) fails to account for primitive tokens that represent uninstantiated properties at all, or there is “too much” asymmetric dependence and we get the contradiction of mutual asymmetric dependence (or a new disjunction problem).

Here is the first alternative. As long as neither unicorn nor shunicorn is instantiated in the actual world, we have no basis for the needed asymmetric dependence: the nomic relation between shunicorn and S’s U-tokens is not asymmetrically dependent on the nomic relation between unicorn and S’s U-tokens. In this case, none of the three versions of (R) can account for S’s U-tokens representing unicorn, and similarly for any other un instantiated property.

Here is the other alternative. If we accept Fodor’s gloss on asymmetric dependence, we get a contradiction. Fodor has explained that “it can be true that the property of being a unicorn is nomologically linked with the property of being a cause of ‘unicorn’-tokens even if there aren’t any unicorns.” He continues:

Maybe this cashes out into something like “there wouldn’t be nonunicorn-caused ‘unicorn’ tokens but that unicorns would cause ‘unicorn’ tokens if there were any unicorns.” And maybe that cashes out into something like this: there are non-unicorn-caused ‘unicorn’ tokens only in nearby worlds in which there are unicorn-caused ‘unicorn’ tokens. But [he adds] . . . I am not an enthusiast for such translations.

_(TC, 46)_

If we took this seriously – though the qualification at the end suggests that we don’t have to – we would say, by parity of reasoning, that there are shunicorn-caused U-tokens only in nearby worlds in which there are unicorn-caused U-tokens. But as Fodor says, there are unicrn-caused U-tokens only in equally nearby worlds in which there are shunicorn-caused U-tokens. So, we get a
contradiction: shunicorn-caused U-tokens, if there were any, would be asymmetrically dependent on unicorn-caused U-tokens and unicorn-caused U-tokens, if there were any, would be asymmetrically dependent on shunicorn-caused U-tokens. (The contradiction could be avoided by taking S's U-tokens to represent unicorn-or-shunicorn; but this is just the disjunction problem again.)

Since Fodor backs away from his gloss by saying, "I am not an enthusiast for such translations," I am not charging him with a contradiction here. But I do believe that he is on the horns of a dilemma: if the asymmetric dependence condition is satisfied by unicorn, then it is equally satisfied by shunicorn, and we get contradiction or disjunction -- and a straightforward counterexample to the analysis. (Both conditions would be satisfied, but, because of contradiction or disjunction, we can not conclude that S's token represents unicorn.) On the other hand, if the asymmetric dependence condition is not satisfied, the reduction cannot handle uninstantiated properties generally, and we are left with a serious gap. Either way, the reduction fails for unicorns.

Of course, there is an obvious way to avoid the problem: Treat 'unicorn' not as a primitive but as a defined term. Treating 'unicorn' as a nonprimitive term would seem reasonable enough, but it would put an unreasonable constraint on the naturalization project. For it would require that the primitive terms represent only instantiated properties.

Such a restriction of the theory to instantiated properties may be described in either of two ways: it makes instantiation of the relevant properties a necessary condition for asymmetric dependence; alternatively, it abandons the pure-informational version for the local-instantiation version (or perhaps the actual history version). Neither description, I think, would be welcome to Fodor. For both tie the theory to what actually exists, as opposed to nomic relations among properties.

But there seems to be a more serious problem. It is implausible to suppose that whether or not a symbol is (semantically) primitive depends on whether or not the property it represents is instantiated. Let me give three examples of the implausibility.

(a) Consider your Doppelganger in a world that has unicorns. Your Doppelganger has never seen a unicorn but has read about unicorns and has seen pictures of them -- just exactly as you have. On the current suggestion, your Doppelganger's symbol that represents unicorn may be (semantically) primitive, but yours must be an abbreviation of symbols that represent instantiated properties. Prima facie, it would be at least odd for there to be two individuals who have been molecular duplicates all their lives, who have the same wide causal histories, and who have mental symbols that represent the same property, but one of whose symbols is semantically primitive and the other not.9

Moreover, on the Fodorean assumption that syntax is "in the head," you and your Doppelganger must have the same syntactically primitive predicates. Thus, if Fodor's view allows you and your Doppelganger to differ in your semantically primitive predicates, the view severs the connection between syntax and semantics. And Fodor would be mistaken in saying that his view explicates "the semantical relation between a syntactically primitive predicate and the property it expresses" (TC, 70; my emphasis).
(b) Consider a would-be natural kind, a property that was widely but mistakenly thought by experts to be instantiated – say, phlogiston. On the current suggestion that "concepts that express uninstantiated properties are ipso facto constructions out of concepts that express instantiated properties" (TC, 67), the concept that expresses phlogiston is ipso facto a construction out of concepts that express instantiated properties. What might those constituent concepts be? The supposition that the concept that expresses phlogiston is "a construction out of concepts that express instantiated properties" does not seem to capture what scientists intended when they postulated phlogiston as a natural kind to explain combustion; nor does it seem to capture what we mean when we say that eighteenth-century scientists incorrectly postulated phlogiston.

(c) Consider symbols that represent artifacts. There was a time when, say, shoe was uninstantiated. On the requirement that only instantiated properties can be expressed by primitive symbols, it would seem that the symbol for shoe (or for any other artifact) could not be primitive – at least before the advent of shoes. And it seems unlikely that, at the first shoe instantiation, the derived symbol for shoe was transformed into a primitive symbol. By like reasoning, no symbol that represented an artifact could be primitive.

Thus, I do not think that representation of uninstantiated properties has been handled satisfactorily. Uninstantiated properties do not seem expressible by primitive symbols, and the alternative view that they are expressible by nonprimitive symbols has the implausible consequences canvassed in (a)–(c).

The Disjunction Problem: Cats/Robot-Cats

Suppose that young Sally – let us keep her isolated from spoken language for a while – lives in an environment populated not only with ordinary cats, but with an equal number of robot-cats. The robot-cats are distinguishable from the cats by knowledgeable people (though not by Sally).10 Sally has seen 1,001 robot-cats, each of which has produced in her a token of nonsemantic type F, and she has never tokened F under any other circumstances. Then, one day, she sees for the first time a cat, which also produces in her a token of type F. Until the cat-caused F-token in question, Sally has never seen or otherwise been exposed to cats, and, of course, we are making no assumptions about what she has been told. (In intentional terms, Sally at this point can not distinguish robot-cats from cats.) What does the cat-caused F-token represent?

There are three candidates. (i) The cat-caused F-token correctly represents a cat, and the other F-tokens have misrepresented robots as cats all along; (ii) the cat-caused F-token misrepresents a cat as a robot; (iii) the cat-caused F-token correctly represents the cat as a cat-or-robot-cat. Fodor opts for the third answer. Let us see why.

(i) Suppose that the cat-caused F-token correctly represents a cat, and that the other F-tokens have misrepresented robot-cats as cats all along. This option is not plausible on the asymmetric dependence view. For it would describe all the robot-cat-caused F-tokens as asymmetrically dependent on the cat-caused F-tokens. But if there is any asymmetric dependence, it is the other way around: The
cat/F connection would be asymmetrically dependent on the robot-cat/F connection.

(ii) Suppose that the cat-caused F-token represents robot-cat, and thus misrepresents the cat as a robot-cat. But this option ignores the relevant counterfactuals. If Sally had encountered cats, they would have caused F-tokens. It is only an accident that she encountered only robot-cats, and Fodor says that "the semantically relevant samples include not just the ones that were encountered, but also the ones that would have been encountered but for an accident." Notice that this response implicitly appeals, not to the pure-informational version but to the local-instantiation version. Thus, instead of saying that the cat-caused F-token represents robot-cat, Fodor (TC; 49) takes the third option.

(iii) Suppose that the cat-caused F-token represents robot-cat-or-cat. Fodor points out

It is OK for some predicates to be disjunctive as long as not all of them are. One can perfectly consistently hold, on the one hand, that "cat" means robot or cat when it's accidental that you learned it just from robot-cats; while denying, on the other hand, that it would mean cat or robot if you had learned it in a world where all you could have learned it from where robot-cats (e.g., because there aren't any cats around). (TC, 49)

First, notice that this just abandons the pure-informational version in favor of the local-instantiation version of the reduction. It is obvious that the local-instantiation version precludes Fodor's favored approach to unicorns.

Second, of course, it is "OK for some predicates to be disjunctive as long as not all of them are." But the cat/robot case is generalizable. For example, suppose that Sally has also seen 1,001 mules, each of which has produced in her an M-token; then one day, for the first time, she sees a horse, which also produces in Sally an M-token. By parity of reasoning, we should say that Sally's first M-token represents mule-or-horse. The same story could be told about (almost?) any symbol. 13

For these reasons, the disjunctive option (iii) seems a risky resting place for Fodor. But let us press on. Let's call the story up to this point 'scene 1.' When we leave scene 1, Sally can not misrepresent a cat as a robot-cat, because her F-tokens represent some disjunctive property cat-or-robot-cat-or perhaps something else.

On to scene 2. Even though robot-cats and cats and perhaps other things initially are in the extension of F-tokens, surely it must be possible at some stage for Sally to misrepresent a cat as a robot-cat. Perhaps, over time, Sally sees lots of cats as well as robot-cats, and observing them closely, comes to respond to cats and robot-cats differently. No longer do cats and robot-cats indiscriminately cause F-tokens, but now cats and robot-cats cause tokens of different types.

Then, one day, Sally sees a cat, which, under the circumstances that day, she misrepresents as a robot-cat. There are three points to notice.

1 The first point concerns the impossibility of error in scene 1. Suppose that from her vantage point in scene 2, Sally thinks back to the time when she saw her first cat (scene 1) and thinks, I mistook that cat for a robot-cat. But if in scene 1, her cat-caused token represented, not robot-cat, but cat-or-robot-cat, as Fodor has
claimed, then there was no mistake to be made. Sally's mistake is to think that she made a mistake.

Fodor replies by offering an "easy answer" and an "interesting answer." The easy answer is that "her indiscriminate application of the same term to both cat and robot-cat was a symptom of her failure to distinguish between them. Not distinguishing between [them] was a serious mistake (by [Sally's] current lights)" (TC: 50). But this reply is not really to the point.

The point is that, in scene 1, the inclusion of both cats and robots in the extension of Sally's F-tokens is not correctly described as a failure or a mistake at all. For anyone who, like Fodor, endorses the disjunctive option (iii), the difference between scene 1 and scene 2 is that Sally's cat-caused tokens represent different properties – cat-or-robot-cat in scene 1 and cat in scene 2. The fact that, in scene 2, Sally knows the difference between cats and robot-cats does not make the cat-caused tokening of F in scene 1 a misrepresentation, when in scene 1, Sally's F-tokens represented cat-or-robot-cat.

What Fodor calls "the interesting answer" requires distinguishing between what is in the extension of the symbol (cats and robot-cats) and what concept a person uses the symbol to express. Fodor describes Sally as having made a mistake, which he describes by saying that "she took it that the robots that she called 'cats' had a certain non-disjunctive property which they shared with everything else in the set (cats U robots). By her present lights, by contrast, there is no such property" (TC, 50).

Perhaps, in scene 2, when Sally thinks back to her days of not distinguishing between cats and robot-cats, she is using a symbol to express a certain property; but in scene 1, there is no question of her using a symbol to express a property, nor of her intending to apply a symbol in a certain way. In scene 1, we are trying to establish what property the primitive symbol represents – solely on the basis of causal relations and asymmetric dependence. Moreover, Fodor's atomism requires that it be possible for Sally to be capable of being in a single intentional state, without attributing to her any other intentional states, such as an intention to use a symbol in one way rather than another.14,15

It seems inconsistent to say both that the "first 'cat' token means cat or robot and is thus true of the cat that it's applied to" (TC, 49) and that its tokening involves a mistake. So, the first point – that Sally's thought in scene 2 that she had mistaken a cat for a robot-cat (in scene 1) was itself a mistake, on Fodor's view – still stands. And this leads to the second point:

2. The second point is that the asymmetric dependence account leaves a large hole. How do we describe, in nonintentional and nonsemantic terms, the change from scene 1 (with no misrepresentation) to scene 2 (with misrepresentation of a cat as a robot-cat). To answer in terms of Sally's ability to distinguish cats from robot-cats does not suffice.

If she is presented with two robot-cats and a cat, Sally can discriminate in the sense of classifying the robot-cats as more like each other than either is like a cat. But this ability can not be the relevant difference between scene 1 and scene 2, for three reasons. First, it is an ability that Sally has in scene 1, as well as in scene 2. Secondly, the ability to distinguish between robot-cats and cats does not suffice for the difference anyway. If Sally is presented with two chihuahuas and a Doberman pincher, she can distinguish in the sense of classifying the chihuahuas
as more like each other than either is like the Doberman; but the three are all
dogs for all that. And if she is presented with two small cats and a large cat, Sally
can classify the two small cats as more like each other than either is to the large
cat, but that does not count against their all being cats. Thus, the ability to
distinguish (put as unintentionally as I can put it) does not account for the
difference between scene 1 and scene 2.

A direct appeal to asymmetric dependence does not explain the change from
scene 1 to scene 2, because the same question re-arises: how do we get from the
absence of asymmetric dependence in scene 1 to its presence in scene 2? In scene
1, the law ‘cat --> F’ was not asymmetrically dependent on the law ‘robot-cat
--> F’ since cats and robot-cats were both in the extension of F-tokens. But if
asymmetric dependence is supposed to account for misrepresentation, and if Sally
can ever misrepresent a cat as a robot-cat (and surely she can in scene 2), then we
need an account of how ‘cat --> F’ comes to be asymmetrically dependent on
‘robot-cat --> F’ for Sally.

3 The third point is that the distinction between scene 1 and scene 2 bears an
uncanny resemblance to the distinction on which Fodor has argued that
teleological theories of content rest, and against which Fodor has argued. The
distinction to which teleological theories seem committed is a distinction between
type 1 situations and type 2 situations. In type 1 situations, “if Ps cause S-tokens,
then S means P (and if P is disjunctive, then so be it)” (TC, 13). In type 2
situations, S-tokens may be caused by nonPs, and they still mean P. Misrepresentation,
according to these theories, is possible only in type 2 situations – just
as, in the story of Sally, her tokens could misrepresent a cat as a robot-cat only in
scene 2.

One difference between Fodor’s theory and teleological theories is this:
according to teleological theories, in type 1 situations, which are “normal” or
“optimal,” only the property in the extension of T causes T-tokens. But since
Fodor does not want to appeal to normal conditions it seems that for him, in type
1 situations (as in scene 1), any property that, but for an accident, would have
caused T-tokens must be included in the extension of T.16

If scenes 1 and 2 of Sally’s story correspond to type 1 and type 2 situations of
teleological theories of content, then Fodor is saddled with the same problems
that he has detected in the teleological theories of content.

For these three reasons, I conclude that Fodor’s reduction has not given an
adequate account of the cat/robot-cat case.

Semantically Relevant Relations

With respect both to the unicorn case and to the cat/robot case, we have seen the
persistent threat of the disjunction problem. Fodor’s solution to that problem lies
in the requirement of asymmetric dependence. In order to make the crucial
asymmetric dependence relation work in the required way, however, Fodor must
restrict its application to semantically relevant relations. He says, “If there’s going
to be a causal theory of content, there has to be some way of picking out
semantically relevant causal relations from all the other kinds of causal relations
that the tokens of a symbol can enter into." He does this by adding a condition to the reduction: robustness.

In this section, I shall first show why robustness, or some other requirement that restricts the asymmetric dependence condition, is needed; then, I shall show why it is unavailable on a physicalistic account.

Fodor appeals to robustness to block counterexamples to his reduction. Suppose that 'A \rightarrow B' is a law and 'B \rightarrow C' is a law. Then the law 'A \rightarrow C' is asymmetrically dependent on the law 'B \rightarrow C'. "Since causal chains give rise to a species of asymmetrical dependence, and since every event belongs to some causal chain or other, how are we to avoid concluding that everything means something? Pansemanticism gone mad." The solution is that "content requires not just causal dependence but robustness too." In the causal chain example, all A-caused Cs are also B-caused, and conversely. "So the asymmetric dependence of 'A \rightarrow C' on 'B \rightarrow C' doesn't satisfy the conditions on robustness; so it's not semantically relevant" (TC: 62).

Let me try to make more vivid the need for a condition (like robustness) in addition to the nomic-relations condition and the asymmetric dependence condition as originally stated. (The differences among the three versions of the reduction are irrelevant to the present point.) Suppose that it is a law that A \rightarrow T and it is a law that C \rightarrow T, where A and C are properties, and T is a nonsemantic type of mental token, and the law 'C \rightarrow T' is asymmetrically dependent on the law 'A \rightarrow T'. Construe asymmetric dependence as Fodor originally defines it: the A/T connection can not be broken without breaking the C/T connection, but the C/T connection can be broken without breaking the A/T connection. Does it follow, as it should from the account, that A is in the extension of T? No. Without some further restriction, it does not even follow that T has content.

Suppose that being at the top of a long escalator causes in Jones feelings of panic, and suppose that being at great heights out in the open causes Jones feelings of panic; and suppose that the escalator/panic connection is asymmetrically dependent on the open-heights/panic connection. Then, the two (original) conditions of nomic relations and asymmetric dependence are satisfied. Yet I do not think that we should say that the feelings of panic represent open heights; indeed, I do not think that the feelings of panic have any intentional object.

Fodor, I believe, would try to block this counterexample by appeal to robustness: the escalator-caused panic is an instance of open-heights-caused panic; hence, all such feelings of panic are nomically related to open heights. If so, there are no non-open-height-caused feelings of panic, but robustness (deployed as a requirement) demands that Ts represent As only if nonAs can cause Ts. So, "if we stipulate that asymmetric dependence engenders content only if it produces robustness," the counterexample may be claimed to fail on the grounds that the asymmetric dependence does not produce robustness.

Originally, robustness is described as a pretheoretical fact about meaning with which the asymmetric dependence condition is claimed to be compatible: "satisfaction of the asymmetric dependence condition is compatible with any amount of heterogeneity in the causal history of 'cow' tokens" (TC: 38). But by
the end of “A Theory of Content,” robustness is deployed, as in the above example, not as an observation about meaning, but as a requirement on the theory: “‘X’ means X only if you can have X-tokens that aren’t caused by Xs” (TC: 71).

It is one thing to say that asymmetric dependence is compatible with causal heterogeneity, another to say that it produces content only when there is causal heterogeneity. Without robustness as a restriction on the asymmetric dependence condition, rather than an interesting fact with which asymmetric dependence is compatible, the theory is subject to counterexamples. But if robustness is taken to be a requirement on the theory (rather than a pretheoretical fact about meaning), and if the theory is to be physicalistic, then robustness must be construed nonsemantically and nonintentionally. Can it be?

As Fodor introduces the idea of robustness, he describes it as an unreduced semantic fact: “‘cow’ tokens get caused in all sorts of ways, and they all mean cow for all that” (TC: 37). This formulation requires that the tokens at issue are typed semantically (as ‘cow’-tokens, not as #c’o’w’#-tokens). Since the point of the theory is to provide a physicalistic reduction, it can not be a condition on the reduction that the relevant tokens already be typed semantically, on pain of circularity. Can robustness be construed nonsemantically, in a way that would allow it to be a requirement of a physicalistic theory?

If so, then the robustness requirement must type X-tokens nonsemantically. But in that case, the robustness requirement is no requirement at all. It rules out nothing. For example, a token of nonsemantic type C, which is usually caused by cows, say, may be produced by an electrical probe. In that case, tokens of that nonsemantic type can be caused by noncows, and they pass the robustness requirement. Since, presumably, tokens of (almost?) any nonsemantic type can be produced by appropriate electrical probes in the brain, as well as by instances of distal properties, tokens typed nonsemantically are always robust.

Indeed, it is less than clear that a nonsemantic construal can do justice to the original intuition behind the idea of robustness. If we take robustness to require that X-tokens can be caused by non-Xs and if we take X-tokens to be a nonsemantic typing (by shape, say), then we get some formulation bordering on the ridiculous: S-shaped-tokens can be caused by non-S-shapes. Thus, as a nonsemantically specified requirement, robustness does not provide the needed restriction on asymmetric dependence. But as an unreduced semantically specified requirement, robustness robs the theory of its physicalistic credentials.

The point that I am trying to make here is subtle. There is no problem with taking robustness to be a semantic fact with which the asymmetric dependence condition is compatible. The difficulty arises in taking robustness to be a part of the theory, a way to rule out irrelevant cases of asymmetric dependence. If robustness (as an unreduced semantic notion) is a condition on asymmetric dependence, then it is not the case that “you can say what asymmetric dependence is without resort to intentional or semantic idiom” (TC: 38). And Fodor’s suggestion that “intentionality equals information plus robustness” (TC: 71) fails to meet the demands of physicalism.

So, robustness can not serve to restrict the asymmetric dependence relations to the semantically relevant ones. If construed semantically, it is question-
begging; if construed nonsemantically, it does not rule out semantically irrelevant relations. Indeed, no requirement can rule out semantically irrelevant relations unless it can give a nonsemantic specification of which relations are semantically relevant. For my part, I doubt that nonsemantic specification of semantic relevance will ever be produced.

Let me venture a diagnosis of what has happened. I think that there has been a conflation of two distinct questions, one of which identifies mental tokens nonsemantically, and the other of which identifies mental tokens semantically, and that this conflation has allowed robustness to slip into the account, unnoticed, as a semantic idea:

(a) Given that token t is a T_n-token (i.e., has a certain nonsemantic narrow property), what determines its extension?  
(b) Given that token t is a 'cat'-token (i.e., has the wide property of representing cat), how can it have any one of an indefinite number of causes and still represent cat?

(a) asks for a solution to the disjunction problem; (b) asks for an account of robustness. The conflation can be seen when Fodor says, “Solving the disjunction problem and making clear how a symbol’s meaning could be so insensitive to variability in the causes of its tokenings are really two ways of describing the same undertaking” (TC: 37).

The disjunction problem, which makes (a) difficult to answer, arises in the determination of the extension of a nonsemantically identified symbol; the robustness problem arises only after the mental token is typed semantically, only after the extension has been established. Again, the robustness problem is to account for the fact that tokens with a variety of causes can represent the same property – a problem that does not arise until the tokens are typed semantically. Therefore, appeal to robustness can not solve the disjunction problem.

Indeed, no answer to (b) produces an answer to the logically prior (a). To see that (a) is the relevant question, ask: What are the relata of the nomic relation to which representation of a cat, say, is to be reduced? How do we complete the statement: “It is a law that cats cause Xs,” where Xs are tokens of a certain type? Is it that cats are nometrically related to

(i) tokens of a type that represents cats?  
(ii) tokens of a type lawfully caused by cats?  
(iii) tokens of some narrow type T_n?

To say, as the first option has it, “It is a law that cats cause tokens of a type that represents cats” would be, at best, an intermediate step. For there is no reduction if the property nometrically related to cats is that of representing cats. The property of representing cats is itself a semantic property. To say, as the second option has it, “It is a law that cats cause tokens of a type lawfully caused by cats” is obviously trivial. By contrast, the third option would yield, “It is a law that cats cause tokens of type T_n”; clearly, only this latter would yield an informative reduction. Only question (a) asks for an answer in terms of
tokens typed nonsemantically. If the theory is to be physicalistic, it must answer (a).

Conclusion

Fodor has emphasized that he is only giving sufficient conditions for a mental token to represent a property. His claim is that if the counterfactuals that he stipulates were true, then a token would mean so-and-so (TC: 42). Even considered apart from the difficulties that I have raised here, this claim is doubly difficult to evaluate: first, it piles counterfactuals on counterfactuals in ways that are not intuitively obvious; and second, there are no clear examples in which the allegedly sufficient conditions do hold. (My intuitions, at least, boggle at comparing nomologically impossible worlds, as the asymmetric dependence condition requires.)

Nevertheless, let me conclude by venturing an outright counterexample (one that does not raise problems of disjunction or robustness) to Fodor's allegedly sufficient conditions for representing a property. Suppose that it is a law that tomatoes cause A-tokens, and that it is a law that apples cause A-tokens, and that the "tomato" law is asymmetrically dependent on the "apple" law. Then, both the conditions of the reduction are satisfied, and by reasoning exactly like that in the cow/cat case, Fodor should say that A-tokens represent apple. But suppose that the "apple" law and the "tomato" law are both asymmetrically dependent on the law that fruit causes A-tokens, and that apples are closer to a paradigm of fruit than are tomatoes. In this case, we should say that A-tokens represent fruit, not apple. That is, even if it is a law that apples cause A-tokens and tomatoes would not cause A-tokens unless apples did, it does not follow that the apple-caused token represents apple.

Whether this toss-away counterexample succeeds or not, let me sum up what I hope to have shown: the disjunction problem, first identified by Fodor himself, is both deep and pervasive, and the account of representation in terms of asymmetric dependence has not solved it.

Along the way, I have argued for the following: (i) Fodor vacillates between two interpretations of the reduction – the pure-informational version (when he considers unicorns) and the local-instantiation version (when he considers cats and robots); (ii) none of the versions of Fodor's view can adequately handle representation of uninstantiated properties generally; (iii) none of the possible options in the cat/robot case is adequate; (iv) Fodor's view collapses into the kind of two-types-of-situation view characteristic of the teleological theories that he criticizes.

In the face of these problems, I conclude that the claim that mental representation has been naturalized has not – or at least not yet – been sustained.

NOTES

1 As we shall see, there really are two distinct problems here – the disjunction problem
and the problem that Fodor (1987d) calls the problem of “robustness” – that are not clearly distinguished. See section “Semantically Relevant Relations.”

2 Fodor (unpublished-c) relativizes his account to persons. In that case, we could say that an Earthian’s XYZ-caused W-tokens (if there were any) would be asymmetrically dependent on her H₂O-caused W-tokens; and a Twin’s H₂O-caused W-tokens (if there were any) would be asymmetrically dependent on her XYZ-caused W-tokens. Since this seems to yield a peculiar idea of what a law of nature is, and since none of my present points turns on relativizing the view or not, I shall ignore the relativization (which seems to me implausible anyway).

3 This is a paraphrase from Fodor’s A Theory of Content (1990b). Hereafter, I shall abbreviate the title by TC and put citations in the text. I wish to thank Fodor for making this important paper, as well as the July 1988 draft, available to me.

4 Officially, asymmetric dependence is a relation between laws, but I shall follow Fodor and speak of Q-caused T-tokens as being asymmetrically dependent on P-caused T-tokens when there are laws “Q → T” and “P → T” such that the former is asymmetrically dependent on the latter.

5 TC, 29. For purposes of reduction, I shall identify the relevant mental tokens nonsemantically. Fodor recognizes the necessity of nonsemantic and nonintentional specification of the tokens whose interpretations are at issue, but he prefers to speak of “cat-caused ‘cow’-tokens,” instead of, say, “cat-caused C-tokens,” where C is a nonsemantic type. (I try to use the expression “‘cow’-tokens” consistently to pick out a semantic type, the type of tokens that represent cow.) I believe that Fodor’s (1987d) equivocal use of expressions like “‘cow’-tokens” – sometimes as a semantic specification and sometimes as a nonsemantic specification – has obscured the difference between the disjunction problem and the robustness problem. See the section “Semantically Relevant Relations.”

6 The term “pure-informational” is Fodor’s.

7 Fodor formulates the actual-history version “if only as an exercise” (TC, 63), in order to show how the asymmetric dependence view can avoid the kind of verificationism that attends the treatment of Twin-Earth cases by the pure-informational view.

8 Since spoken languages like English stand as much (or as little) in need of naturalization as thought, no facts about the semantics of natural language can help in naturalizing mental representation (unless one has an independent account that naturalizes English). On Fodor’s view, the semantics of spoken language derive from the semantics of mental representation. So, specification of the semantics of primitive mental symbols must not appeal to the fact that such-and-such is a word in English, nor even to the fact (if it is one) that a thinker is a speaker of a public language.

9 On the actual-history version, as opposed to the local-instantiation version, a causal theorist could respond that “unicorn” could not be semantically primitive for my Doppelganger any more than it could be for me since neither of us has had any causal interactions with unicorns. I thank Gabriel Segal for pointing this out to me.

10 This case was originally in Baker, 1989. Fodor’s response is in TC, pp. 48–50. I am tightening up the example here in two ways: I am making it explicit that what property the P-token represents must be determined by facts free of assumptions about spoken language; and I am making it explicit that the robots look like a species of cats that knowledgeable people can distinguish from real cats. In the earlier version, this latter point was implicit since Sally herself later comes to know that the robot-cats are not cats.

11 TC, 50; draft of July 1988. Fodor makes the same point, but less clearly, I think, in the March 1989 draft.

12 Fodor also says that the cat/robot case is underdescribed, because I do not say whether or not Sally has a standing intention to use ‘cat’ as a kind term (TC, 60). But attribution
of any such intentions at this point would be way ahead of the game. The issue here
is not how a thinker uses (or intends to use) a symbol, but how we theorists are to
interpret it (without assuming that the thinker knows a spoken language). At this
fundamental level, the interpretation of primitive Mentalese symbols must be
understood solely in terms of causal relations and asymmetric dependence.

Moreover, the disjunctive option (iii) seems to compromise robustness. As we shall see
in the next section, the matter of robustness is not straightforward. To see that there
are two formulations of robustness, compare TC, 68 with TC, 62.

I am not assuming that Sally has any intentions about natural kinds. (I think it
empirically unlikely that children distinguish natural kind terms in Mentalese or in
spoken language from others. In learning any new term, the learner has the intention
of “going on in the same way” — whether the new term is ‘robot-cat’, ‘basket’, ‘break’,
‘promise’, or anything else.) Also, it seems unlikely that Sally was capable of
distinguishing disjunctive from non-disjunctive properties in scene 1. If we must
impute to a thinker sophisticated cognitive maneuvering in order to interpret
Mentalese primitives, then it seems unlikely that the naturalization project will ever get
off the ground.

Throughout his account, and especially when dealing with Twin-earth cases, Fodor
attributes to the thinker settled policies or intentions to apply a term in a certain way.
His reason for thinking that such attribution is not question-begging has two parts:
First, what matters is the truth of certain counterfactuals — e.g., that in a world where
there are both water and XYZ and they are distinguishable, an Earthian’s T-w-tokens
would track H₂O and th. Twin-Earthian’s T-w-tokens would track XYZ — and it is the
intentions that make the relevant counterfactuals true. Second, “semantical relations
hold in virtue of counterfactuals which can be specified naturalistically by quantifying
over (e.g., intentional) mechanisms in virtue of whose operations the counterfactuals
hold.” (See Fodor (unpublished-c), comments on my APA paper, “What is a Mental
Representation?”, 9 (Baker 1988).
But the question is this: if you, the theorist, do not have access to the content of the
intentions, how do you know over which mechanisms to quantify? But if you do have
access to the content of the relevant intentions, then you must assume that there is a
naturalistic account of that content. But you cannot make such an assumption in the
context of offering a naturalistic account, on pain of circularity.

As Fodor puts it in the July 1988 draft: “the extension of S’s term T includes things
of type X whenever (i) no Xs have been encountered by S, (ii) if S had encountered
Xs, S would have tokened T; and (iii) S’s failure to encounter Xs was fortuitous” (TC:
50; July 1988).

TC: 38; emphasis his. I think that the expression “tokens of a symbol” in this
quotation displays the equivocation that I shall identify below, cf. note 20.

I believe that there are some typographical errors in Fodor’s statement of the example
in the draft of the text that I have.

At any rate, as we have just seen, Fodor responds to other putative counterexamples
in this way. Put aside for a moment suspicion that the story does satisfy the robustness
requirement: if you trick the subject into thinking that he is at a great open height, but
he is not, his same feelings of panic are caused by non-open-heights. As we shall see,
robustness cannot do the job assigned to it anyway.

This use of “cow token,” thought crucial for the point about robustness, violates
Fodor’s reply to Block’s objection. (See TC: 56–8.) From Psychosemantics on, there is
an ambiguity in Fodor’s use of expressions like “‘cat’-token” or “‘A’-token,”
expressions employing single quotes. Sometimes he takes ‘cat’-tokens as being typed
by what they represent — as when he speaks of ‘red’-tokens as tokens that represent red.
But when he first entertains the disjunction problem, he considers whether ‘A’-tokens
represent As and misrepresent Bs or represent the disjunctive property (A or B) (Fodor, 1987d: 101). In this latter usage, ‘A’-tokens are not typed by what they represent.

Also, the conception of law in play here needs a good deal of spelling out.

This description of the apple case parallels Fodor’s description of the cow/cat case. However, I am not confident that Fodor’s two conditions can ever be applied nonquestion-beggingly. It seems to me at least arguable that our calling (in public language) apples and tomatoes both fruit is the reason that apples and tomatoes both cause ‘fruit’-tokens (if they do), and not the other way around.

The assumption that apples are closer to being a paradigm of fruit that are tomatoes is needed to make the counterfactuals work out right: The tomato/A connection is asymmetrically dependent on the apple/A connection in that tomatoes wouldn’t cause ‘fruit’ tokens if apples didn’t. The apple/A connection is asymmetrically dependent on the fruit/A connection in that apples wouldn’t cause ‘fruit’ tokens if fruit generally didn’t.

I read earlier versions of this paper to helpfully critical audiences at Brandeis University, Duke University, the University of Nebraska (Lincoln), and SUNY at Buffalo, as well as at the 1988 Eastern Division meeting of the American Philosophical Association, where Jerry Fodor responded. I especially want to thank Gabriel Segal for commenting on a penultimate draft. Work on this paper has been generously supported by the Woodrow Wilson International Center for Scholars and by Middlebury College.