# Intentionality & Naturalism[1]

...the deepest motivation for intentional irrealism derives not from such relatively technical worries about individualism and holism as we've been considering, but rather from a certain ontological intuition: that there is no place for intentional categories in a physicalistic view of the world; that the intentional can't be *naturalized.*
     (Fodor, 1987, p. 97)

## 1. Catastrophe Theory

Intentional irrealism is the doctrine that meaning is a myth. A bit more precisely, it is the claim that nothing in the world instantiates intentional properties — that intentional predicates are true of nothing. If intentional irrealism is correct, then it is not the case that

(1) 'Snow is white' means that snow is white.

or that

(2) George Bush often thinks about winning the next election.

or that

(3) Lincoln wanted to free the slaves.

Nor is it the case that

(4) Thinking about winning the election sometimes causes Bush to smile.

or that

(5) Lincoln's desire to free the slaves caused him to sign the Emancipation Proclamation.

     Obviously, intentional irrealism has some very startling consequences. If it is true then a very substantial part of what we read in our textbooks, teach our children and say to each other is mistaken. Indeed, as Fodor has remarked, with only a bit of hyperbole, if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying ..., if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world.[2]
     Though we rather doubt that the world would come to an end, perhaps Fodor is closer to the mark in claiming that if intentional irrealism is correct and commonsense intentional psychology really were to collapse, that would be, beyond

---

[1] We are grateful to Brian McLaughlin for many hours of helpful discussion. Earlier versions of this paper were presented at CUNY Graduate Center, Notre Dame University, and the University of North Carolina - Greensboro. Comments and criticisms from those audiences are acknowledged with thanks.

[2] Fodor (1990a), p. 156.

comparison, the greatest intellectual catastrophe in the history of our species.... The collapse of the supernatural didn't compare.[3]

Very well, then, let's agree that intentional irrealism is a very radical doctrine. But why on earth should anyone *worry* about it? Why does anyone think it is even remotely plausible? In the quote with which we began this paper Fodor maintains that the ''deepest motivation for intentional realism'' is the suspicion ''that the intentional can't be naturalized.'' Viewed as a bit of sociology, it's our guess that Fodor is right. In recent years, many philosophers have put a very high priority on providing a ''naturalistic'' account of intentional categories.[4] Moreover, there is an unmistakable tone of urgency in much of this literature. Naturalizing the intentional isn't just an interesting project, it is vitally important. *Something dreadful* will follow if it doesn't succeed. And for many writers, we suspect, that dreadful consequence is intentional irrealism.[5] But this sociological fact raises a philosophical puzzle. *Why* would irrealism (or some comparably unsettling conclusion) follow if ''the intentional can't be naturalized?'' What is the connection between the existence or non-existence of a naturalistic account of intentional categories and the truth or falsehood of claims like (1) - (5)? These are the questions that motivate this paper.

To answer them, of course, it will be necessary to say just what is involved in ''naturalizing'' the intentional. And, as we shall see, there is no shortage of answers to choose from. But not just any answer will do. A satisfactory account of what it is to ''naturalize the intentional'' — an account that makes sense of what Fodor sees as ''the deepest motivation for intentional irrealism'' — will have to satisfy a pair of constraints. First, it will have to sustain an argument from the premise that intentional notions can't be naturalized to the conclusion that intentional irrealism or some other deeply troubling doctrine is true. Second, there must be some reason to think that, when ''naturalizing'' is unpacked along the lines proposed, it is in fact the case that the intentional can't be naturalized. For even if non-naturalizability entails irrealism, this is surely nothing to

---

[3] Fodor (1987), p. xii.

[4] See, for example, Block (1986), Devitt (1990), Dretske (1981, 1988), Field (1978), Fodor (1984, 1987, 1990a, 1990b), Loar (1981), Lycan (1988), Millikan (1984), Papineau (1987), Schiffer (1982), Stalnaker (1984).

[5] Schiffer provides a characteristically forthright illustration of this attitude. On his view, the question of how the semantic and the psychological are related to the physical is ''an urgent question'' since ''we should not be prepared to maintain that there *are* semantical or psychological facts unless we are prepared to maintain that such facts are completely determined by, are nothing over and above, physical facts.'' (1982), p. 119.

On Fodor's view, the urgency of the issue reaches to the very core of contemporary academic life. For if the intentional can't be naturalized, then lots of people who work in cognitive science should no longer get government sponsored research grants.

If it turns out that the physicalization — naturalization — of intentional science ... is impossible, ... then it seems to me that what you ought to do is do your science in some other way.... [I]f you really can't give an account of the role of the intentional in the physical world .... [then] by Christ... we should stop spending the taxpayer's money. (Fodor, 1990c), pp. 202-203.

worry about if the claim that the intentional can't be naturalized is neither intuitively plausible nor supported by a convincing argument.

It is our contention that, while various accounts will satisfy one or the other of these constraints, there is no account of what it is to naturalize the intentional that will satisfy *both* of them. To support our contention, we will survey a number of proposals on what ''naturalization'' comes to and we will go on to argue that none of these candidates will satisfy both of the constraints. Obviously this strategy won't provide a conclusive case for our conclusion, since there may be some quite different account of naturalizing that does satisfy both constraints. But if so, we haven't a clue about what it might be.

If we are right, if there is no account that satisfies both constraints, then there is something deeply misguided about the urgency that imbues so much of the recent literature in this area. It may, of course, be perfectly reasonable to adopt one or another account of what it would be to naturalize the intentional, and to explore the possibility of bringing it off. A successful naturalization might well be an impressive and valuable accomplishment. But if it should turn out that intentional notions can't be naturalized, *no dire consequences will follow*. We will not have to rewrite history, or renounce intentional psychology, or revise the way we describe and explain people's behavior. It will not be the end of the world. It won't even be the beginning of the end.

Before launching into our survey of accounts of ''naturalizing'' a few words are in order on some of the other troubling consequences that might be thought to follow if naturalization does not succeed. In Fodor's writing, and elsewhere in the literature, the dominant worry is the one that has been center stage in this section: if the intentional can't be naturalized, then intentional irrealism will have won the day. But often enough one finds suggestions of other calamities that may ensue if naturalization fails. One of these is that intentional states might turn out to be causally impotent. In the passage quoted earlier, for example, Fodor frets that it's the end of the world if it isn't literally true that his wanting is causally responsible for his reaching, and his believing is causally responsible for his saying. In Fred Dretske's writing the worry that intentional states might turn out to be causally inert is frequently cited as a motive for seeking a naturalized account of these states. Indeed, Dretske sometimes suggests that if intentional states are causally impotent, then perhaps we should not include them in our ontology at all.

> If beliefs and desires are not causally relevant to behavior, I, for one, fail to see why it would be worth having them.... If reasons aren't causes, one of the chief — indeed (for certain people) the *only* — motive for including them in one's inventory of the mind, vanishes.[6]

Another, rather different concern is that if naturalization fails, then there could be no serious *science* of intentional psychology because there could be no *laws* that invoke intentional terms or intentional properties. We are no more impressed by these worries than we are about the concern over irrealism. For, as we shall argue in the sections that follow, on any reading of the claim that the intentional can't be naturalized which is even remotely likely to be true, neither of these calamitous consequences would follow.

---

[6] Dretske (1989), p. 1. See also Dretske (1988), p. 80, and Dretske (1990), p. 6.

## 2. Naturalizing and Conceptual Analysis

Once upon a time something called ''conceptual analysis'' was all the rage in philosophy. The journals, back then, were filled with attempts to provide necessary and sufficient conditions for the application of a term or a concept. And, more often than not, when one philosopher published such an ''analysis'' another philosopher would describe a hypothetical situation in which we would intuitively say that the analysans applied and the analysandum did not, or vice versa. For people who remember those bygone days (only one of us does), much of the literature on naturalizing the intentional provokes a strong sense of *deja vu*. Consider, for example, the following quote:

> The worry about representation is above all that the semantic (and/or the intentional) will prove permanently recalcitrant to integration in the natural order.... What is required to relieve the worry is therefore, at a minimum, the framing of *naturalistic* conditions for representation. That is, what we want at a minimum is something of the form *'R represents S' is true iff C* where the vocabulary in which condition C is couched contains neither intentional nor semantic expressions.[7]

Of course, an interest in providing necessary and sufficient conditions is not, by itself, enough to convict a philosopher of engaging in conceptual analysis. For typically a conceptual analyst will not be happy with just any set of conditions that happen to be co-extensive with the predicate being analyzed. If a proposed analysis is to be acceptable, it has to be the case that the co-extension obtains not only in all actual cases, but in imaginary or hypothetical cases as well. The bi-conditional specifying the analysis must not only be true, it must be *necessary*. Moreover, the alleged co-extension in all possible worlds is supposed to be testable by consulting our linguistic intuition and determining what we would say about hypothetical cases. This method would seem to make the most sense if we suppose that the co-extension derives from the meaning of the concepts that underlie our predicates — the analysans (the right-hand side of the bi-conditional) unpacks the meaning of the concept expressed by the analysandum.

Is it the case that Fodor and others who worry about the possibility that the intentional can't be naturalized are actually worried about the possibility that the meaning of intentional predicates or intentional concepts can't be set out as a set of necessary and sufficient conditions which do not themselves invoke intentional terms? We're not at all sure. Indeed, it is our suspicion that these philosophers have *no* clear idea of what ''naturalizing'' amounts to, and that much of their anxiety can be traced to this confusion. But if it is not clear that these philosophers really want a conceptual analysis, it is clear that if ''naturalizing'' is understood in this way, it will not satisfy the first of our two constraints.

---

[7] Fodor (1984), p. 32.

Indeed it is rather ironic that Fodor often seems to be troubled by the fact that our intentional concepts can't be analyzed in non-intentional terms. For among contemporary philosophers no one has been more adamant than Fodor in insisting that we should not expect our terms or concepts to be analyzable *at all*. Here is an example of the sorts of things he says when this mood is upon him:

> [I]t seems to me to be among the most important findings of philosophical and psychological research over the last several hundred years (say, since Locke first made the reductionist program explicit) that attempts at conceptual analysis practically always fail.
>
> Consider, for example, the failure of the reductionist program within the study of language.... [W]hat I'll call the *Definition Hypothesis* [is the claim that] (a, weak version) ... many de facto lexical concepts are definable; and (b, strong version) that they are definable in a vocabulary of sensory-terms-plus-logical-syntax.
>
> It's simply notorious that the stronger version of this claim has proved to be untenable.... But what's equally true, and considerably more striking, is that the evidence seems to bear against the definition hypothesis even in the weak version; if there are no plausible cases of definition in a sensory vocabulary, there are also remarkably few plausible examples of definition in a *non*-sensory vocabulary, one indication of which is the striking paucity of working examples in the standard literature. There is 'bachelor', which is supposed to mean 'unmarried man'; ... there are jargon terms, which are explicitly and stipulatively defined; ... there is a handful of terms which belong to real, honest-to-God axiomatic systems; ... and then there are the other half million or so items that the OED lists. About these last apparently nothing much can be done.[8]

On our view, there can be no serious quarrel with Fodor's assessment of the track record of conceptual analysis. Though lots of very clever people have tried very hard to produce them over the centuries, we still have no plausible definitions for 'knowledge' or 'cause' or 'law' or 'freedom,' or for any of the other terms that loom large in philosophical discussion. Moreover, as Fodor goes on to illustrate, it is no easier to provide definitions for more mundane terms like 'paint' or 'parent' or 'pig'. The more one plays the game of trying to provide exceptionless, intuitively acceptable necessary and sufficient conditions, the more one is inclined to accept Fodor's conclusion: ''[W]hen it comes to definitions, the examples almost always don't work.''[9]

What are we to make of this situation? Well, of course, it might be that conceptual analysis is just *hard*, and that if we keep at it we will ultimately succeed in producing a significant number of intuitively acceptable definitions. However, it is also entirely possible that we will never succeed — that the project of defining most common predicates is simply impossible.

---

[8] Fodor (1981), pp. 283-4.

[9] Fodor (1981), p. 288.

If it *is* impossible, this will have important consequences for those parts of philosophy and psychology that deal with the structure of human concepts. There is a venerable tradition in this area which assumes that the concept or mental structure underlying the use of most predicates is actually a mentally represented definition — a set of necessary and sufficient conditions. In deciding whether or not a term applies to a given case, this "Classical View" maintains, we are either consciously or (more typically) unconsciously determining whether the case at hand satisfies the conditions of the definition. If it turns out that there just are no definitions for most terms, then obviously the Classical account of the structure and use of concepts will have to go.[10]

In recent years there has been a growing realization that the Classical account of concepts is in deep trouble, and a number of interesting alternatives have been proposed. Perhaps the best known of these are the prototype and exemplar accounts of concepts developed by Eleanor Rosch and her associates. On the prototype theory, concepts are weighted lists of features that are characteristic of the most typical members of the category that the concept picks out. The list will generally include lots of features that are not necessary for category membership. On the exemplar story concepts are, in effect, detailed mental descriptions of particular members of the category. Thus, for example, the concept underlying your use of the word 'dog' might include a detailed descriptions of Lassie and Rin Tin Tin. In determining whether to categorize something as a dog, this theory maintains, you assess the similarity between the target and the various exemplars stored in semantic memory.[11] Fodor has proposed a very different alternative to the Classical account of concepts. On his view, the concepts that underlie most of our one-word predicates have no structure at all — or at least none that is relevant to the semantic properties of the concept. Of course if this is right it is very hard to see how these concepts might be learned. And that's just fine with Fodor, since he thinks they are all innate.[12]

This is not the place to elaborate the details of these various "non-Classical" theories of concepts, or to debate their virtues and shortcomings. Our reason for mentioning them was simply to make clear that there are lots of interesting theories about concepts on the market which are compatible with (and which might well explain) the finding that most of our concepts appear to have no intuitively acceptable definitions. So if it is indeed the case that most concepts have no definitions, there is nothing much to worry about. Rather, the appropriate response is to get busy and try to determine which of the various non-Classical theories of concepts is correct. It would, by contrast, be simply *mad* to think that if most of our concepts can't be defined, then the terms that express those concepts are not true of anything. The inference from *The predicate '--- is a pig' cannot be defined* to *There are no such things as pigs* is simply perverse. Concern about porcine irrealism is not even a remotely appropriate reaction to the

---

[10] For a useful discussion of the Classical View, see Smith & Medin (1981), Chapter 3.

[11] See Smith & Medin, Chapters 4-6.

[12] For Fodor's view see Fodor (1981).

collapse of the Classical theory of concepts. But, of course, exactly the same can be said about intentional predicates. Perhaps there are good reasons to worry about intentional irrealism being true, but the fact that 'R represents C' can't be defined surely isn't one of them.[13]

What about the other two concerns that we sketched at the end of Section 1?. If intentional terms can't be defined, does it follow that intentional states are causally impotent, or that there are no laws invoking intentional properties? In both cases, we maintain, the answer is clearly *no*. To see why, consider a few analogies. If the Classical theory of concepts is wrong, then there will be no way to provide necessary and sufficient conditions for predicates like 'x shot y' or 'z died'. But from this, surely, it does not even begin to follow that it is not literally true that being shot by John Wilkes Booth caused Lincoln to die. And, of course, if the Classical theory is wrong, then terms

---

[13] In the passage from Fodor (1984) quoted at the beginning of this section, he insists that ''what we want at a minimum is something of the form *'R represents S' is true iff C....*'' But in later papers Fodor is prepared to accept a lot less. In the following passage, for example, he no longer insists on necessary and sufficient conditions. Rather, he tells us, merely sufficient conditions will do.

> I want a naturalized theory of meaning; a theory that articulates, in nonsemantic and nonintentional terms, sufficient conditions for one bit of the world to be about (to express, represent, or be true of) another bit. (Fodor (1987) p. 98; see also Fodor (1990), pp. 51-52.)

But as noted by Jones, Mulaire & Stich (1991), if we read him literally this is just too easy. Here are two sufficient conditions that seem to meet Fodor's requirement.

> If R is Fodor's most recent utterance of ''Meaning Holism is a crazy doctrine'' (or the thought that underlies it) then R is *about* Meaning Holism, and R is true iff Meaning Holism is a crazy doctrine.

> If R is Laurence's most recent utterance of ''Madonna is daring'' (or the thought that underlies it) then R is about Madonna and expresses the proposition that Madonna is daring.

Obviously, it would be an easy task to produce indefinitely many more. But perhaps this reading is uncharitably literal. Perhaps what Fodor requires in a naturalized theory of meaning are sufficient conditions which follow from the meaning of the terms involved. It is easy enough to provide intuitively plausible sufficient conditions of this sort for many non-intentional terms. Here's one:

> For all x, if x is a sow, then x is a pig.

But, of course, examples like this are cheating. In the spirit of Fodor's requirement that sufficient conditions for representation or aboutness be stated in *nonsemantic* and *nonintentional* terms, we should require that the sufficient conditions for being a pig be stated in *nonporcine* terms. Once *this* requirement is imposed, however, providing meaning based sufficient conditions for being a pig looks to be just about as intractable as providing a full blown definition. If it is impossible to provide such sufficient conditions, that will be an interesting result in lexical semantics. But it will *not* entail that there are no pigs. Similarly, if it turns out that meaning based sufficient conditions cannot be given for intentional locutions, it will not follow that meaning is a myth.

like 'force,' 'mass,' and 'gravity' won't be definable either. But it would be at best a bad joke to conclude, from this, that there are no laws that invoke these terms. If the Classical view of concepts collapses, it will not take all of physics with it. The situation seems entirely parallel for intentional terms. If it turns out that they can't be analyzed or defined, this would provide no reason at all to conclude that intentional states are causally impotent, or that there are no laws invoking them. So if ''naturalizing the intentional'' requires providing a Classical analysis of intentional concepts, then if the intentional can't be naturalized, we have found no reason to think that anything at all troublesome will follow.

## 3. Naturalizing, Natural Kinds and Essential Properties

To set the stage for our second account of what it might be to naturalize the intentional, we'll begin with a brief reminder of some very influential doctrines in the philosophy of language. Consider so-called ''natural kind'' predicates like 'water' or 'gold'. What is it that determines which parts of the world are in the extension of such predicates? According to the widely discussed causal/historical account of reference, the answer to this question must invoke the notion of ''essential properties'' of natural kinds — properties that everything in the extension of a natural kind term must have. A bit fancifully, the causal/historical story might be sketched as follows:

> A kind term first acquires its referent when it is used to ''baptize'' or ''dub'' some newly noted samples of the stuff to which the term will refer. This process is sometimes described as ''grounding'' the predicate. Once the predicate has been grounded, it can be transmitted from one speaker to another in appropriate communicative settings. And those to whom the predicate is passed can pass it on again. The speakers who originally ground the predicate need have no deep understanding of the nature of the stuff they are dubbing; indeed, they may have all sorts of wildly mistaken beliefs about it. The speakers who acquire the predicate via reference preserving transmissions need never have come in contact with anything in the extension of the predicate. They too can harbor many false beliefs about the nature of the stuff to which the term refers.

Now obviously there is something missing in this tale. For a predicate like 'gold' gets grounded on just a few samples of gold. And yet the extension of the predicate must include *all* the gold that ever has or ever will exist in the universe. What is the relation between the dubbed samples and the rest of the gold in the universe, in virtue of which the dubbing succeeds in attaching the term to all gold, wherever it may be? It is here that the doctrine of *essential properties* is typically brought into play. The basic idea is that individual items are grouped into natural kinds in virtue of the possession of certain essential properties, and it is the job of science to discover what these properties are. Thus, for example, science tells us that having atomic number 79 is the essential property of gold, that being $H_2O$ is the essential property of water, and so on. When a natural kind term gets grounded, the term comes to apply not only to the samples present at the dubbing but also to everything else in the universe that has the same essential properties.

How does all of this relate to the project of naturalizing the intentional? To see the answer, let's go back to the quote from Fodor near the beginning of the previous section. What was worrying Fodor was that intentional categories might ''prove permanently recalcitrant to integration in the natural order.'' And what was required to relieve the worry was ''a framing of naturalistic conditions for representation ... something of the form *'R represents S' is true iff C* where the vocabulary in which condition C is couched contains neither intentional nor semantic expressions.'' Our first pass at unpacking this requirement was to view it as a demand for a conceptual analysis. But it could equally well be viewed as asking for a specification of an underlying essential property — the property in virtue of which the predicate 'R represents S' applies to all and only those pairs of things in the universe such that the first represents the second. On this interpretation, the bi-conditional needed to naturalize the representation relation would have a status akin to the one Putnam and others have attributed to bi-conditionals like:

(6)  (x) x is water iff x is $H_2O$.

It is a necessary truth, but its necessity has nothing to do with the structure of the concept that speakers invoke when they use the terms involved. It isn't known *a priori*, and it can't be discovered by probing intuitions or by doing psycholinguistics. The only way to discover it is to do the appropriate sort of science.

How likely is it that *this* is what philosophers want when they set about trying to naturalize the intentional. Well, there are some practitioners of the craft who offer accounts of representation that rely heavily on notions borrowed from science (typically evolutionary biology). Some of these writers go out of their way to explain that they are not trying to capture our intuitions about representation, and thus are not worried by the fact that their analyses have counter-intuitive consequences.[14] All of this is compatible with the interpretation that these philosophers are seeking an account of the essential properties of representation. But we don't propose to press the point since, as we noted earlier, we rather suspect that most of the writers who worry about naturalizing the intentional have no clear idea of what ''naturalizing'' amounts to. What is clear is that if ''naturalizing'' is interpreted in this way, then once again it will not satisfy our first constraint.

One way of arguing for this claim would be to mount a head-on assault on the whole idea of scientifically discoverable essential properties, and on the account of the reference of natural kind terms that goes along with it. There is already a substantial literature pointing out the shortcomings of this rather trendy package of ideas, and we have considerable sympathy with the emerging critique.[15] But all that would make a very long argument, and we have a much shorter one to offer.

Suppose it is the case the doctrine of essential properties and the associated story about reference can survive serious scrutiny. Suppose further that when ''naturalizing''

---

[14] See, for example, Millikan (1989), pp. 290-291.

[15] Canfield (1983), Donnellan (1983), Leplin (1979, 1988), Shapere (1982).

is interpreted in the way we've just sketched, it turns out that 'R represents S' and other intentional predicates cannot be naturalized. Would this be enough to make intentional irrealism plausible? Surely the answer is *no*. To see the point, we need only note that there are endlessly many predicates for which no one would even dream of seeking scientifically discoverable essential properties. Yet it would be simply perverse to claim that these predicates can't be truly applied to anything. Nobody seriously thinks that anything remotely analogous to (6) will be available for such one place predicates as 'couch,' 'car,' 'war,' 'famine,' or 'die,' nor for two place predicates like 'owns,' 'kills,' 'throws,' 'mates with,' 'fixes,' or 'crushes'. But it would be preposterous to suggest that this entails there is no killing or war or famine, and that no one ever owns anything or dies. If natural kind terms are defined as those whose extension is determined by scientifically discoverable essential properties, then one way of putting our point is that there are many, many predicates that are not natural kind terms, and the fact that they are not natural kind terms is no reason at all to suppose that they cannot be truly predicated of anything. So if it turns out that nothing analogous to (6) is forthcoming for intentional predicates, the right conclusion is not that those predicates are true of nothing, but simply that, in the sense lately defined, they are not natural kind terms. And that would hardly be the end of the world.

Could it be that while intentional irrealism doesn't follow from the fact that intentional predicates aren't natural kind terms, something comparably unsettling does follow? Let's take a brief look at the pair of possibilities suggested at the end of Section 1. The first of them focuses on the causal efficacy of intentional states. Might it be the case that if intentional predicates aren't natural kind terms in the sense we've defined, then they can't be used to make causal claims that are literally true. This strikes us as a singularly implausible suggestion. For, as we noted earlier, it is literally true that being shot by John Wilkes Booth caused Abraham Lincoln to die, though neither 'shoots' nor 'dies' is likely to be the sort of term whose extension is determined by scientifically discoverable essential properties. So even if it turns out that intentional predicates are not natural kind terms in the sense we've defined, the causal efficacy of intentional states and processes might still be on a par with the causal efficacy of shooting, crushing, eating or mating. And that should be efficacy enough for anyone.

A second possibility is that if intentional predicates aren't natural kind terms, then perhaps there could be no science of intentional psychology. For, it might be argued, such a science would have to include intentional laws, and laws can only be stated with natural kind terms. No kind terms, no laws; no laws, no science. Now as we see it, the problem here comes with the link between kind terms and laws. *Why* can laws only be stated with natural kind terms? One might view it as simply a stipulative definition: natural kind terms just are the sorts of terms that can occur in lawlike statements. But now we have a potential equivocation on our hands. For we have been assuming that natural kind terms are defined as those whose extension is determined by scientifically discoverable essential properties, and the current argument proposes a very different definition. Of course it might be claimed that these two definitions pick out the same class of terms — that all and only terms whose extension is determined by scientifically discoverable essential properties can be used in lawlike statements. But we find this a singularly

implausible proposal. For in sciences far removed from psychology there appear to be lots of terms invoked in laws for which nothing much like (6) is in the offing. We see no reason at all to suppose there are scientifically discoverable essential properties that fix the reference of terms like 'inflation,' 'fitness,' 'mass,' 'gravity,' or 'electric charge,' for example. If this is right — if there are lots of terms invoked in scientific laws whose extensions are not fixed in the way that the causal/historical theory claims the extensions of terms like 'gold' are fixed — then the putative threat to intentional psychology disappears.

Thus far we have been arguing that an account of naturalizing the intentional which requires producing something akin to (6) will not satisfy the first of our two constraints. Neither intentional irrealism nor any other catastrophic consequence follows if the intentional can't be naturalized, when naturalizing is interpreted in this way. But we are also inclined to think that if we take seriously the story about reference that serves as a backdrop for the current proposal on naturalizing, then our second constraint will not be satisfied either. For if that story is correct, then the usual arguments aimed at showing that the intentional can't be naturalized just don't go through.

Those arguments typically begin by describing some feature or cluster of features that are important or essential for intentional states, *on the commonsense account of these states*. The arguments then try to show that respectable scientific theories cannot accommodate states with the features in question. The conclusions the arguments draw are just the ones that Fodor feared: that the intentional ''will prove permanently recalcitrant to integration in the natural order,'' and that ''there is no place for intentional categories in a physicalistic view of the world.'' However, if the causal/historical account of reference is correct, then the conclusions of these arguments do not follow from the premises. For on the causal/historical account, the essential properties that determine the extension of natural kind terms are to be discovered by science, and our commonsense views about the things we are referring to with natural kind terms may be wildly, hopelessly wrong. Indeed, the fact that ignorance and error do not undermine reference is taken to be a major selling point of the causal/historical theory.[16] But if our commonsense views about the things we are referring to may be seriously mistaken, then the (alleged) fact that commonsense imbues intentional states with scientifically unacceptable features entails nothing at all about the scientific respectability of intentional states. For commonsense may just be *wrong*; our intentional terms may actually refer to states that do not have these scientifically unacceptable features. So if the causal/historical theory of reference is correct, there can be no serious argument from premises about the commonsense characterization of intentional states to conclusions about role that the intentional states refered to by commonsense psychology might play in scientifically acceptable theories. Without some argument along those lines, however, it is hard to see why we would have any reason to believe that the intentional can't be naturalized.[17]

---

[16] See, for example, Devitt & Sterelny (1987), Secs. 4.2 & 5.2.

[17] For some elaboration on the argument set out in the last two paragraphs see Stich (1991a; in prep.)

## 4. Naturalizing and Supervenience

> It's hard to see ... how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe of their supervenience on?) properties that are themselves *neither* intentional *nor* semantic. If aboutness is real, it must be really something else.[18]

### 4.1. The Game Plan

Thus far we haven't done very well in finding interpretations of ''naturalizing'' that satisfy our two constraints. But in the passage just quoted, Fodor seems to be making a pair of suggestions that we haven't yet explored. To avoid irrealism, intentional properties must be *identical with* or *supervene upon* non-intentional properties. So perhaps naturalization should be explained in terms of property identity or supervenience. In the current section we'll consider whether either of *these* proposals satisfies our two constraints. Actually, we will focus almost entirely on supervenience, since on all plausible accounts of that notion, it is a weaker relation than identity. Indeed, on most accounts, property identity entails supervenience, and thus non-supervenience entails non-identity. So if nothing nasty follows from the fact that the intentional doesn't supervene on the non-intentional, then the fact that intentional properties are not identical with non-intentional ones will be no cause for worry.

In restricting our attention to supervenience we are not exactly making things easy for ourselves, however. For the literature on supervenience has blossomed profusely during the last few years, and this literature suggests a variety of different ways in which the idea that the intentional supervenes on the non-intentional may be spelled out.[19] These alternatives differ on a pair of dimensions. First, the notion of one class of properties supervening on another can be explicated in two different ways, one of which (so-called *strong supervenience*) entails the other (*weak supervenience*). Second, there are various options that might be proposed as the ''supervenience base'' for intentional properties — the class of properties on which intentional properties are expected to supervene. In the arguments that follow we will restrict our attention to weak supervenience. For, since strong supervenience entails weak supervenience, the failure of weak supervenience entails the failure of strong. Thus if we can show that no untoward consequences follow when weak supervenience does not obtain, the same conclusion will follow if strong supervenience fails.

Here's the game plan. We'll begin with a brief explanation of the two notions of supervenience. We'll then attend to three different candidates that might be proposed as the supervenience base for intentional properties. In each of these three cases we will

---

[18] Fodor (1987), p. 98.

[19] Haugeland (1982), Horgan (1982), Kim (1978, 1982, 1984, 1987), Lewis (1983), Petrie (1987), Teller (1984).

argue that the constraints set out in Section 1 are not met. In the first two cases, it is the first constraint that isn't satisfied: Neither irrealism nor the other unwelcome consequences follow if supervenience fails. In the third case, it is the second constraint that isn't satisfied. For in this case it is wildly implausible that supervenience fails. We will follow all of this with a brief discussion of another notion of supervienience, so called *global supervenience*, whose precise relation to the other two notions is a matter of some dispute. Here again, we will argue, nothing catastrophic follows if intentional properties fail to supervene on the various bases that have been proposed. End of game plan. It's time to get to work.

## 4.2. Two Notions of Supervenience

Supervenience is usually construed as a relation between two classes of properties. So to begin, let us adopt the following convention. Let B and S be two classes of properties (think of them as the Base class and the Supervenient class) whose members are $b_1$, $b_2$, ..., $b_i$,... and $s_1$, $s_2$, ... $s_i$,... respectively. Now the basic idea is that one class of properties, S, supervenes on a second, B, if the presence or absence of properties in the first class is completely determined by the presence or absence of properties in the second class. There are various ways in which this basic idea can be made more precise.

Perhaps the most intuitive way to proceed is to exploit the notion of a B- or S-doppelgänger. A B-doppelgänger of an object is an object that has exactly the same B properties as the original. An S-doppelgänger is one which has exactly the same S properties. Thus, for example, if B includes only two sorts of properties, height and weight, then your B-doppelgängers are all and only those things that have the same height and weight that you do. One vivid way to explicate the various versions of the idea that B properties determine S properties is to use the picturesque language of possible worlds. If in all possible worlds, every pair of B-doppelgängers that exist in that world are also S-doppelgängers, then we will say that *S weakly supervenes on B*. So if S weakly supervenes on B, then in any possible world we select, if we know that a pair of objects in that world share the same B-properties, we know they share the same S-properties as well. And if a pair of objects in that world do not share the same S properties, we know that there must be at least one B property that one has and the other doesn't. We can build a stronger notion of supervenience if we relax the restriction that the B-doppelgängers are in the same world. We will say that *S strongly supervenes on B* if all B-doppelgängers of an object, no matter what possible world they inhabit, are also S-doppelgängers. Obviously, strong supervenience entails weak supervenience. Plainly there are lots of other distinctions that might be drawn by restricting attention to one or another special class of possible worlds. But we will leave all of that to the aficionados. Henceforth, when we use 'supervenience' we will mean weak supervenience, as characterized above, unless otherwise specified.

## 4.3. The Supervenience Base: Three Proposals

On what sorts of properties might it be thought (or hoped) that intentional properties should supervene? As we read the literature, there are at least three proposals for the Base class on which intentional properties must supervene if nasty consequences are to be avoided. We propose to consider each of these proposals, proceeding from the most restrictive to the least.

**4.3.1.** The first idea is that something untoward will follow if the intentional properties of an organism do not supervene on the *current, internal, physical properties* of the organism. These are the properties that organisms share with their Putnamian doppelgängers — the hypothetical particle for particle replicas that exist in some far corner of space-time.[20] And if intentional properties supervene on current internal physical properties then in any given world, organisms must have the same intentional properties as their Putnamian doppelgängers. What makes this proposal particularly interesting is that it is widely agreed that there are possible worlds in which organisms and their Putnamian doppelgängers *do not* share all of their intentional properties. Indeed, that's the main point that Putnam's famous thought experiment was supposed to establish. George Bush has many beliefs *about* Michail Gorbachev; he has no beliefs at all about Twin-Gorbachev, the atom for atom replica in some far off corner of the universe. The situation is just the opposite for George Bush's doppelgänger. Twin-Bush has lots of beliefs about Twin-Gorbachev, and none about the Gorbachev who leads his life in our part of the universe. But while there is considerable agreement about the fact that at least some intentional properties don't supervene on current, internal physical properties, there is much less agreement on what unwelcome consequences this failure of supervenience is supposed to entail. Let's consider the options.

First on our list, as always, must be the specter of intentional irrealism — the thesis that intentional properties aren't ''real properties of things.'' But surely intentional irrealism would be a preposterous conclusion to draw from the fact that intentional properties don't supervene on current, internal physical properties. For there are *lots* of properties of objects that don't supervene on their current, internal physical properties — often, it would appear, for much the same reason that intentional properties do not. And it would be quite absurd to suggest that non-supervenience entails irrealism in *all* these cases. To see the point, consider a few examples. There are lots of copies of Picasso paintings in the world. And some of them are astoundingly accurate. Let us imagine that someone produces a ''perfect'' copy — a canvas that is an atom for atom duplicate of the original. Of course, the perfect copy would still be a *copy*, it wouldn't be the original. For to be an original Picasso, a canvas must have the right *history* — it must actually have been painted by Picasso. Much the same point can be made about real $100 bills. A master counterfeiter might produce a bill that is an atom for atom replica

---

[20] This idea, or something like it, is suggested in Stich (1978, 1983) and in Fodor (1980, 1987, 1991). For the original account of Putnamian doppelgängers, see Putnam (1975).

of one produced by the Bureau of Engraving and Printing. But it would still not be a *real* $100 bill. Indeed, as Fodor has noted, not even *God* can make a real $100 bill. Only a branch of the U.S. Treasury can do that.[21]  It follows, then, that neither the property of being an original Picasso nor the property of being a genuine $100 bill supervenes on the current, internal physical states of an object. So if a property's failure to supervene on current, internal physical states were sufficient to show that nothing has the property, then it would follow that there are no genuine Picassos or real $100 bills. But that, of course, is just silly. The idea that intentional irrealism follows from failure to supervene on current, internal physical states is equally silly.

Before pushing on, it will be useful to mention a rather different sort of example. Both genuine Picassos and real $100 bills are artefacts. And it might be thought that natural properties or categories are not linked to history in this way. But this is almost certainly a mistake. To see the point, consider the classification of organisms into species. Regardless of how similar a pair of organisms are, it is plausible to suppose that they will not count as members of the same species unless they also share the appropriate sort of evolutionary history. If there are creatures in Australia that evolved from birds, then they do not count as members of the same species as Stich's cat, Eggplant, no matter how similar their current, internal physical states and Eggplant's current internal physical states may be. And if scientific explorers on Mars should come upon a macro-molecule that is an atom for atom replica of an HIV virus isolated on Earth, it would not be an HIV virus unless it shared a common evolutionary ancestry with HIV viruses found on Earth. If this is right, then the property of being a cat and the property of being an HIV virus do not supervene on the current, internal states of the entities that have those properties. But here again, it would be simply absurd to conclude that there are no such things as cats or HIV viruses.

What we have been arguing in the last two paragraphs is that intentional irrealism does not even begin to follow from the fact that intentional properties do not supervene on the current, internal physical states of organisms. Let's now ask whether one of our other discomforting conclusions follows from the failure of the intentional to supervene on this sort of Base. Does it, perhaps, follow that intentional states, though they exist, must be causally impotent — that believings can't cause sayings, that wantings can't cause scratchings — and thus that the end of the world is near? It seems clear that the answer is no. To see why, consider an analogy. Suppose some poor fellow, call him Henry, is crushed to death when an original Picasso sculpture falls on him. Being crushed by an original Picasso caused Henry to die. In some possible world in which Henry exists, we may suppose that he has a Twin who is crushed to death by an atom for atom identical statue, but one which was not made by Picasso. So being crushed by an original Picasso does *not* cause Twin-Henry to die. Nonetheless it is ''literally true'' that being crushed by an original Picasso caused Henry to die. Consider now the case of intentional causation. Suppose that both Bush and Twin-Bush say, ''Gorbachev is bold.'' Only Bush believes that Gorbachev is bold, however; Twin-Bush believes that

---

[21] Fodor (1987), p. 45.

Twin-Gorbachev is bold. Does this difference somehow entail that it could not be ''literally true'' that Bush's belief caused his utterance? Since it appears that this case is entirely parallel to the previous one, it's hard to see why we should be skeptical about one causal claim and not about the other.

Another worry that one might have at this point focuses on the causal efficacy of *properties* rather than *states*. The concern might be put like this:

> Though it is true enough that Bush's belief state causes his utterance despite it's failure to supervene on his current, internal, physical properties, it isn't true that this state causes his utterance *in virtue of being the belief that Gorbachev is bold.* What is worrisome about this sort of failure to ''naturalize'' the intentional is that it makes intentional properties causally irrelevant.[22]

Now we are none too clear about how one goes about determining the causal efficacy of properties. But, for argument's sake, let us grant that if intentional properties do not supervene on the current, internal, physical properties of organisms, then intentional properties are not causally efficacious. Would this be a major catastrophe? So far as we can see, it would be no catastrophe at all. For given any intentional property, it is easy to find a ''narrow'' surrogate of that property which *does* supervene on the current, internal, physical state of the organism. Following Stich (1991b), we can take the property of believing that [p] to be the narrow surrogate of believing that p. The extension of the expression, ''____ believes that [p]'' is just the class of all possible individuals who believe that p along with all of their current-internal-physical-property doppelgängers.[23] Similarly, we could construct a ''narrow'' surrogate for the property of being an HIV virus. The extension of this surrogate property would be the class of all possible entities that have the property of being an HIV virus, along with all their current-internal-physical-property doppelgängers. Here, again, the narrow surrogate *will* supervene on the current, internal physical states of the entities in question. Thus even if we grant that intentional properties (and properties like being an HIV virus) are not causally efficacious, there is no reason to fear that the end of the world is near. In both cases, the properties fail to be causally efficacious because they have historical or relational components ''built in''. But it is easy enough to characterize narrow surrogates that factor out the historical or relational components. And we see no reason at all to suppose that these narrow surrogates are not causally efficacious. It's hard to think that even Fodor's Granny could ask for more.[24]

---

[22] For a vivid illustration of this worry, see Dretske (1988), pp. 79-80.

[23] Alternatively, if the reference to possible individuals is problematic, we can take the extension of ''____ believes that [p]'' in a given possible world to be the class of all individuals in that world who believe that p, and all their current-internal-physical-property doppelgängers in that world, and all individuals in that world who are current-internal-physical property doppelgängers of individuals in other possible worlds who believe that p. (The account in Stich (1991b) neglects this last conjunct.)

[24] For more Fodor's Granny and her views, see Fodor (1987), passim, and Loewer & Rey (1991), p. ii.

Let's turn to the worry about laws. Does the fact that intentional properties don't supervene on current, internal physical states indicate that they cannot play a role in laws? There are, in the literature, a number of arguments aimed at establishing some sort of link between laws and properties that supervene on the current, internal physical states of systems. But we don't propose to tackle these arguments head-on, for, if the truth be known, we are not at all sure we really understand them.[25] But we are sure that when one starts looking at cases, the proposed link seems very implausible. Consider the HIV virus. Though the details are still to be worked out, it is plausible to assume there is a lawlike connection between infection by the HIV virus and the death of certain cells that play an important role in the immune system. Thus something like the following might well turn out to be a law:

> For all x, if x is infected by the HIV virus (and certain further conditions are met), then most of x's T-cells will be destroyed.

But if the current worry were correct, then there could be no such law, because being infected by the HIV virus is not a property that supervenes on an organism's current, internal physical state. For a rather different example, consider Greshem's Law which claims that bad money drives good money out of circulation. Plainly, neither the property of being money nor the properties of being good and bad money supervene on the current, internal, physical state of coins, banknotes, wampum and the like. But this is no reason at all to suppose that Greshem's law is mistaken. Analogously, the fact that intentional properties do not supervene on the current, internal physical states of organisms does not entail that intentional properties cannot play a role in laws. So the reading of ''naturalizing the intentional'' which requires showing that the intentional supervenes on the current, internal, physical state of the organism, fails to satisfy our first constraint. If the intentional can't be naturalized (in this sense) nothing on our list of unwelcome consequences will follow.

**4.3.2**. A second proposal for a supervenience base widens the base class by dropping the restriction to *current*, *internal* states. On this proposal, intentional properties will be naturalized if we can show that they supervene on *physical* properties of the organism. Though it is not entirely clear which properties to count as physical properties, a natural way to construe the current proposal is to take the physical properties to be those that might be invoked in physical laws. When the proposal is construed in this way, however, just about everything we said in the previous section can be repeated with minor modifications. More specifically:

i) Intentional properties do not supervene on physical properties. The crucial difference between Bush and Twin-Bush is that the former has had appropriate causal interactions with Gorbachev, while the latter has had completely parallel interactions with

---

[25] One of these arguments is to be found in Fodor (1987), Chapter 2. Another is to be found in Fodor (1991). Fodor himself no longer claims to understand the first of these. For a critique of the second, see Christensen (forthcoming).

Twin-Gorby. But having had appropriate causal interactions with Gorby (rather than Twin-Gorby) is not the sort of property that is likely to be invoked in a physical law.

ii) Being a genuine Picasso or a real $100 bill don't supervene on physical properties either. Thus

iii) If the fact that a property, P, does not supervene on physical properties is sufficient to establish that nothing has P, then we would have to be irrealists about genuine Picassos and real $100 bills. And that's absurd.

iv) So the failure of the intentional to supervene on the physical will give no support at all to intentional irrealism.

v) The properties that differentiate people who are crushed by original Picassos from their Twins who are crushed by perfect copies are not properties that will be invoked in physical laws. But it may still be literally true that Henry's death was caused by being crushed by an original Picasso. Analogously, the fact that Bush believes that Gorbachev is bold, while Twin-Bush believes that Twin-Gorbachev is bold, does not entail that Bush's utterance was not caused by his belief. And finally,

vi) The properties that distinguish real HIV viruses from their atom for atom duplicates on Mars are not properties that physics is likely to invoke. Nonetheless, it may well turn out to be a law that if a person is infected by HIV, then most of his or her T-cells will die. So the fact that a property does not supervene on physical properties does not preclude it from being invoked in a law. Thus the failure of the intentional to supervene on the physical would not entail that intentional properties can't be invoked in laws.

We conclude that the second proposed base does no better than the first. If naturalizing the intentional means showing that the intentional supervenes on the physical, then if the intentional can't be naturalized, none of our catastrophic consequences will follow.

**4.3.3.** The final proposal for a supervenience base that we will consider is the one that Fodor seems to be urging in the quote with which we began this section. If semantic and intentional properties are real properties of things, he urges, they must be identical with or supervene on ''properties that are themselves neither intentional nor semantic.'' So let's ask whether our constraints are satisfied if we construe naturalizing the intentional to require that intentional properties weakly supervene on the class of all non-intentional and non-semantic properties. The answer we would urge is no. But in this case the problem is with the second constraint not the first. For when naturalizing is understood in this way, the claim that the intentional can't be naturalized is extremely implausible - indeed it may be incoherent. To see the point, we need only remind ourselves of what has to be the case if one class of properties, S, does not supervene on another, B. For supervenience to fail, there must be a possible world in which there are B-doppelgängers that are not S-doppelgängers. That is, there must be objects, x and y, in some world that share all of their B-properties but do not share all of their S-properties. On the current

proposal the B-properties are *all* non-intentional and non-semantic properties. So the B-doppelgängers, x and y, must share their physical properties, their relational properties, their spatial location, their temporal location and their history. But surely if x and y share *all* of these properties, then x and y are *identical*. And if x and y are identical, then they share all their properties, including their intentional properties.

On the current reading of what naturalizing comes to, it would indeed be a catastrophe if the intentional could not be naturalized. For if this happened then in some possible world there would be a single object which both did and did not have a certain property, and logic itself would crumble.[26] Fortunately, there is not the slightest reason to take this prospect seriously.

## 4.4. Global Supervenience

Before bringing this paper to a close, we propose to take a brief look at a third strategy for spelling out the idea of one class of properties supervening on another, the one that goes by the label *global supervenience*. In defining both weak and strong supervenience, the notion of *objects* that were B- or S-doppelgängers of one another played a central role. But, as the name suggests, in global supervenience the central notion is that of *worlds* that are doppelgängers of one another. A pair of possible worlds are doppelgängers of one another with respect to a given property if and only if the total distribution of the property in one of those worlds is the same as the total distribution of the property in the other. So, for example, a possible world which is exactly like our world except for the fact that Stich's cat, Eggplant, has a black nose rather than a pink one would be a shape- and size-doppelgagger of the actual world. But that world would not be a color-doppelgänger of the actual world. With this notion in hand, we can define global supervenience as follows: A class of properties, S, *globally supervenes* on a class of properties, B, if and only if all possible worlds that are B-doppelgängers are also S-doppelgängers. So if S globally supervenes on B, then if a pair of worlds are indistinguishable with respect to the properties in B, they will also be indistinguishable with respect to the properties in S.

In the previous section, we considered three proposals for the base class on which it might be thought that intentional properties should supervene. The first of these, the class of current, internal, physical properties of an object, has no obvious application when global supervenience is at issue. But the other two, the class of physical properties and the class of all non-intentional properties, might both be proposed as a global supervenience base for the class of intentional properties. Let's consider each of them in turn.

---

[26] This argument will not work if naturalizing is unpacked in terms of *strong* supervenience, since in that case the B-doppelgängers might be in different possible worlds. Our view about the account of naturalizing that requires *strong* supervenience on *all* non-intentional properties is much the same as our view — set out at the end of 4.4 — about the account that requires *global* supervenience on those properties.

Recall that, as we proposed to unpack the notion, a physical property is one that might be invoked in a physical law. Do intentional properties globally supervene on physical properties, when physical properties are construed in this way? The answer, we think, is clearly no. For it seems extremely plausible to suppose that there is a possible world, $W_1$, that is a physical doppelgänger of the actual world as it exists right now, but which has no history at all. $W_1$ is one of those worlds that Russell often worried about. It was created just a few seconds ago, fully stocked with phony fossils and light waves racing toward earth just as they would be if they had been emitted by stars millions of years ago. But if $W_1$ has no history, then, according to many philosophers, the distribution of intentional properties in $W_1$ must be very different from the distribution of intentional properties in our world. For in our world, Laurence has lots of beliefs about Julius Caesar; he is connected to Caesar in just the right way to have these beliefs, whatever that way is. But in $W_1$, Laurence has no beliefs about Caesar. There was no Caesar to have beliefs about in $W_1$, so Laurence couldn't be connected to him in the right way.

Very well, then, intentional properties do not globally supervene on physical properties. What follows? Nothing terribly troublesome, so far as we can see. The arguments here are pretty much the same as those in 4.3.2. There are *lots* of properties that do not globally supervene on physical properties — the property of being a genuine Picasso, for example (there are no genuine Picassos in $W_1$), and the property of being a real $100 bill. But from the fact that these properties do not globally supervene on the physical, it surely does not follow that there are no real Picassos and $100 bills in our world. Analogously, from the fact that the intentional doesn't globally supervene on the physical, it does not follow that intentional properties are not instantiated in our world. The property of being an HIV virus doesn't globally supervene on the physical either. But from this we cannot conclude that this property can't be invoked in laws, nor, alas, can we conclude that being infected by HIV doesn't cause people to die. And here, again, the situation for intentional properties looks to be exactly the same.

What about the broader base, the class of all non-intentional properties. Do intentional properties globally supervene on this base? Once again, so far as we can see, the answer is no. For it certainly seems to be logically possible for there to be a world, $W_2$, that is a non-intentional doppelgänger of the actual world, but in which trees or cars or dead people have beliefs or desires or some other intentional states. And it also seems logically possible for there to be a world, $W_3$, that is a non-intentional doppelgänger of the actual world, but in which Dan Quayle has no thoughts at all — he's just a mindless organic robot. The sorts of worlds we are imaging are, near enough, the sorts that some property dualists suppose the actual world to be. And whatever problems one might think this sort of property dualism confronts, it certainly does not seem to be a logically incoherent view. If it is not logically incoherent, if worlds like $W_2$ and $W_3$ really are possible, then intentional properties do not globally supervene on non-intentional properties. But it is hard to see why anyone would think that catastrophic consequences follow. Surely the logical possibility of a world like $W_2$ or $W_3$ does not entail that intentional properties are not instantiated in the actual world. Nor, so far as we can see, does it even begin to entail that *in our world* intentional states are casually impotent or that they cannot be invoked in laws of nature.

At this point, we fear, a resolute opponent might begin fiddling with the notion of *possibility* that is embedded in the definition of global supervenience. Such an opponent might suspect that problems will arise if there are pairs of *nomologically possible* worlds or *metaphysically possible* worlds that are non-intentional doppelgängers but not intentional doppelgängers. The path on which our imagined opponent has embarked is not one we're tempted to follow, for we suspect that it leads directly to a metaphysical swamp. Moreover, even if unwelcome consequences really do follow in these cases — and we see no clear reason to suppose that they do — we are inclined to think that both of them violate our second constraint. It is certainly not *intuitively* plausible that there are pairs of nomologically or metaphysically possible worlds that *are* non-intentional doppelgängers but *are not* intentional doppelgängers. Most people, including many who seem to have exquisitely subtle metaphysical intuition, have no intuitions at all about matters like this. So until someone presents a plausible argument that such world pairs are nomologically or metaphysically possible, we see no reason to take the prospect seriously.

## 5. Conclusion

It's time to sum up. We began with Fodor's observation that ''the deepest motivation for intentional irrealism derives ... from a certain ontological intuition: ... that the intentional can't be *naturalized*.'' But we have had no success at all in making sense of this motivation. If the motivation is to stand up to scrutiny, there must be some account of what naturalizing the intentional comes to which satisfies a pair of constraints. First, the account must sustain an argument from the premise that the intentional can't be naturalized to the conclusion that nothing satisfies intentional properties (or perhaps to the conclusion that intentional states are causally impotent, or to the conclusion that there can be no intentional laws). Second, the claim that the intentional can't be naturalized must not turn out to be utterly implausible. None of the accounts we have been considering satisfy both of these constraints. Of course, it is always possible that there is some other account that will satisfy the constraints. But at this point we think the ball is in the other guy's court. Until some account of naturalizing is given that satisfies both constraints, the most plausible view is that the motivation that Fodor recounts is simply confused. There may be good reasons to take the prospect of intentional irrealism seriously, but the worry that the intentional can't be naturalized is not one of them.

Stephen Stich
Stephen Laurence
*Rutgers University*

# References

Baker, L. (1987). *Saving Belief*, Princeton, Princeton University Press.

Block, N. (1986). ''Advertisement for a Semantics for Psychology,'' in P. French et. al., eds., *Midwest Studies in Philosophy: Studies in the Philosophy of Mind*, Minneapolis, University of Minnesota Press.

Canfield, J. (1983). ''Discovering Essence,'' in C. Ginet & S. Shoemaker, eds., *Knowledge and Mind*, Oxford, Oxford University Press.

Christensen, D. (forthcoming). ''Causal Powers and Conceptual Connections,'' to appear in *Analysis*.

Devitt, M. & Sterelny, K. (1987). *Language and Reality*, Cambridge, MA, Bradford Books / MIT Press.

Devitt, M. (1990). ''A Narrow Representational Theory of the Mind,'' in W. Lycan, ed., *Mind and Cognition*, Oxford, Basil Blackwell.

Donnellan, K. (1983). ''Kripke and Putnam on Natural Kind Terms,'' in C. Ginet & S. Shoemaker, eds., *Knowledge and Mind*, Oxford, Oxford University Press.

Dretske, F. (1981). *Knowledge and the Flow of Information*, Cambridge, MA, Bradford Books / MIT Press.

Dretske, F. (1988). *Explaining Behavior*, Cambridge, MA, Bradford Books / MIT Press.

Dretske, F. (1989). ''Reasons and Causes,'' *Philosophical Perspectives*, 3.

Dretske, F. (1990). ''Does Meaning Matter?'' in E. Villanueva, ed., *Information, Semantics and Epistemology*, Oxford, Basil Blackwell.

Field, H. (1978). ''Mental Representation,'' *Erkenntnis*, 13.

Fodor, J. (1980). ''Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology,'' *Behavioral and Brain Sciences*, 3.

Fodor, J. (1981). ''The Present Status of the Innateness Controversy,'' in J. Fodor, *Representations*, Cambridge, MA, Bradford Books / MIT Press.

Fodor, J. (1984). ''Semantics, Wisconsin Style, *Synthese*, 59. Reprinted in Fodor (1990). [Page references are to Fodor (1990).]

Fodor, J. (1987).  *Psychosemantics*, Cambridge, MA, Bradford Books / MIT Press,

Fodor, J. (1990a).  *A Theory of Content and Other Essays*, Cambridge, MA, Bradford Books / MIT Press.

Fodor, J. (1990b).  ''Psychosemantics, or: Where Do Truth Conditions Come From?'' in W. Lycan, ed., *Mind and Cognition*, Oxford, Basil Blackwell.

Fodor, J. (1990c).  ''Roundtable Discussion,'' in P. Hanson, ed., *Information, Language and Cognition*, Vancouver, University of British Columbia Press.

Fodor, J. (1991).  ''A Modal Argument for Narrow Content, *Journal of Philosophy*, 88.

Haugeland, J. (1982).  ''Weak Supervenience,'' *American Philosophical Quarterly*, 19.

Horgan, T. (1982).  ''Supervenience and Microphysics,'' *Pacific Philosophical Quarterly*, 63.

Jones, T., Mulaire, E. & Stich, S. (1991).  ''Staving Off Catastrophe: A Critical Notice of Jerry Fodor's *Psychosemantics*,'' *Mind and Language*, 6, 1.

Kim, J. (1978).  ''Supervenience and Nomological Incomensurables,'' *American Philosophical Quarterly*, 15.

Kim, J. (1982).  ''Psychophysical Supervenience,'' *Philosophical Studies*, 41.

Kim, J. (1984).  ''Concepts of Supervenience,'' *Philosophy and Phenomenological Research*, 45.

Kim, J. (1987).  '''Strong' and 'Global' Supervenience Revisited,'' *Philosophy and Phenomenological Research*, 48.

Leplin, J. (1979).  ''Reference and Scientific Realism,'' *Studies in History and Philosophy of Science*, 10.

Leplin, J. (1988).  ''Is Essentialism Unscientific?'' *Philosophy of Science*, 55.

Lewis, D. (1983).  ''New Work for a Theory of Universals,'' *Australasian Journal of Philosophy*, 61.

Loar, B. (1981).  *Mind and Meaning*, Cambridge, Cambridge University Press.

Loewer, B. & Rey, G., eds., (1991).  *Meaning in Mind:  Fodor and His Critics*, Oxford, Blackwell.

Lycan, W. (1988).  *Judgement and Justification*, Cambridge, Cambridge University Press.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*, Cambridge, MA, Bradford Books / MIT Press.

Millikan, R. (1989). ''In Defense of Proper Function,'' *Philosophy of Science*, 56.

Papineau, D. (1987). *Reality and Representation*, Oxford, Basil Blackwell.

Petrie, B. (1987). ''Global Supervenience and Reduction,'' *Philosophy and Phenomenological Research*, 48.

Putnam, H. (1975). ''The Meaning of Meaning,'' in K. Gunderson, ed., *Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science* 7.

Schiffer, S. (1982). ''Intention Based Semantics,'' *Notre Dame Journal of Formal Logic*, 23.

Shapere, D. (1982). ''Reason, Reference and the Quest for Knowledge,'' *Philosophy of Science*, 49, 1.

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

Stalnaker, R. (1984). *Inquiry*, Cambridge, MA, Bradford Books / MIT Press.

Stich, S. (1978). ''Autonomous Psychology and the Belief-Desire Thesis,'' *The Monist*, 61.

Stich, S. (1983). *From Folk Psychology to Cognitive Science*, Cambridge, MA, Bradford Books / MIT Press.

Stich, S. (1991a). ''Do True Believers Exist?'' *The Aristotelian Society*, Supplementary Volume LXV.

Stich, S. (1991b). ''Narrow Content Meets Fat Syntax,'' in B. Loewer & G. Rey, eds., *Meaning in Mind: Fodor and His Critics*, Oxford, Blackwell.

Stich, S. (in preparation). *Deconstructing the Mind*.

Teller, P. (1984). ''A Poor Man's Guide to Supervenience and Determination,'' *Southern Journal of Philosophy*, Supplement to Volume 22.