Fodor's New Theory of Content and Computation

ANDREW BROOK AND ROBERT J. STAINTON

Abstract: In his recent book, *The Elm and the Expert*, Fodor attempts to reconcile the computational model of human cognition with information-theoretic semantics, the view that semantic, and mental, content consists of nothing more than causal or nomic relationships, between words and the world, or (roughly) brain states and the world. In this paper, we do not challenge the project. Nor do we show that Fodor has failed to carry it out. Instead, we urge that his analysis, when made explicit, turns out rather differently than he thinks. In particular, in some places where he *sees* problems, he sometimes *shows* that there is no problem. And while he *says* two conceptions of information come to much the same thing, his analysis *shows* that they are very different. This is rather strange.

1. Introduction

Two old friends show up early in Fodor's recent book *The Elm and the Expert* (hereafter *E&E*). First old friend: psychology must employ intentional concepts such as belief and desire. Second old friend: cognitive processes consist of computations and 'computational processes are ones defined over syntactically structured objects' (Fodor, 1994, p. 8). What's new is Fodor's view of intentional content. Narrow content is out; information theoretic semantics is in.[1]

Address for correspondence: Andrew Brook, 2217 Dunton, Carleton University, Ottawa, Canada K1S 5B6.
Email: abrook@ccs.carleton.ca.
[1] Incidentally, Fodor underplays the attractions of narrow content. Far from being some unwelcome stepchild to be embraced only in theoretical extremis, narrow content is an intuitively plausible notion. First, it seems that our contents stay with us no matter what causal environment we find ourselves in. Second, we are aware of many of our mental contents in a way that seems inconsistent with mental content being broad. There seems to be an interesting asymmetry here between mental content and semantic content, however. It is easier to think of the latter as relational than the former. Among other reasons, we often do not know the meaning of a word but it is not easy to think of ourselves not knowing the contents of our thoughts, desires, perceptions, etc.

As we read it, *E&E* is about one central problem brought about by this change: whether the old picture of cognition as a computational process can be made to jibe with the new view that content is information, that content consists in a brain/world relationship. Could computational processes correctly 'track' content thus understood? We do not want to challenge the project. Nor do we aim to show that Fodor has failed to carry it out. Instead, we urge that his analysis, when made explicit, turns out rather differently than he thinks. First we will try to get the problem a little clearer, then we will consider peculiarities in Fodor's proposed resolution of it.

2. The Problem: The Argument for Incompatibility

By *computationalism* we mean the view that psychological states and processes are implemented computationally, where a computation is an operation over syntactic objects: a mapping from symbols to symbols, such that the transformations pay attention only to form, never to content. Let *intentionalism* be the view that psychological states are ineliminably content bearing. Intentionalism can be many different things, depending on how 'intentional' is read. In particular: combining intentionalism with the view that content is information yields what might be called *informational intentionalism*. To avoid such an ugly label, we will speak of 'infointentionalism'. It's the view that psychological states and process are ineliminably intentional, *in the sense of being information bearing*.

With these rough and ready definitions in hand, let us now pose the Question: are computationalism and infointentionalism compatible? Here's an argument, reconstructed from *E&E*, that they are not. We call it the Argument for Incompatibility:

> *Premise One*: If psychological states and processes are ineliminably intentional and psychological states and processes are implemented computationally then there must be computationally sufficient conditions for the instantiation of intentional properties.
> *Premise Two*: Content, being information, is relational.
> *Premise Three*: If content is relational then there aren't any computationally sufficient conditions for the instantiation of intentional properties.

Now, the antecedent of Premise One follows directly from intentionalism, and therefore from infointentionalism, when conjoined with computationalism. So, computationalism plus infointentionalism plus Premise One entail that there *must be* computationally sufficient conditions for the instantiation of intentional properties. Premise Two *also* follows from infointentionalism. And from Premises Two and Three it follows that there *aren't any* computationally sufficient conditions for the instantiation of intentional properties. Evidently, these two conclusions are inconsistent. It begins to look, then, as

though one must give a negative answer to the Question: infointentionalism and computationalism are not consistent.

But there may be hope yet for the view that cognition is a matter of computations which track information. Perhaps one could reject one of the Premises. Premise Two, we repeat, is entailed by infointentionalism. And Premise Three is motivated as follows: computational properties, being syntactic, are internal. But no internal property is such that satisfying it is sufficient for having an external relation. (This is presumably what Fodor has in mind when he remarks that, 'It's as though one's having ears should somehow guarantee that one has siblings.' [Fodor, 1994, p. 14]) Applying this to the case of computational and intentional states/processes, it seems that computational properties cannot possibly guarantee intentional properties—the former are internal while (ex hypothesi) the latter are relational. Hence Premise Three. Given the solidity of Premises Two and Three, then, Fodor goes after Premise One.

The appeal of Premise One resides in our need for what Fodor calls 'a property theory' connecting intentional laws with their computational implementations. As Fodor puts it:

> If the implementing mechanisms for intentional laws are computational, then we need a property theory that provides for *computationally sufficient conditions* for the instantiation of intentional properties. (Fodor, 1994, p.12; his emphasis)

To this Fodor gives a natural reply: the demand is too strong. In fact, he urges, psychological states and processes could be ineliminably intentional *and* be implemented computationally, even if there were no computationally sufficient conditions for the instantiation of intentional properties. If so, then despite the need for the theory to which Fodor refers, Premise One is too strong; indeed, Not True. In sum: the consequent of Premise One is inconsistent with the consequent of Premise Two, and the antecedents of both are true. But Premise One is not true. So, as far as the foregoing argument shows, the answer to the Question may well be 'yes': computationalism and infointentionalism are compatible.

3. *The Revised Argument for Incompatibility*

So far so good. But a still small voice is insistently asking, 'Does the need for a property theory not commit us to *anything*?' Indeed it does. However, Fodor urges, all it commits us to is:

> *Premise One (Revised)*: If psychological states and processes are ineliminably intentional and psychological states and processes are implemented computationally, then the coinstantiation of the com-

putational implementer and intentional implemented must be *reliable*.

Here is the argument that Premise One (Revised) is strong enough to allow for intentional psychological laws. One condition sufficient for computations to track content correctly would be supervenience: all differences of content being reflected in a difference within the computational system, across all possible worlds. That something this strong would *suffice* is presumably what motivates premises like our original Premise One. But, argues Fodor, all we *require* for psychology are computational conditions that reliably link intentional content to syntactic states. Exceptions are quite alright so long as they *are* (just) exceptions, not counterexamples: that is to say, so long as they are infrequent and unsystematic—particularly unsystematic. (Psychology, goes the mantra, is a *ceteris paribus* science, not a basic one.) Moreover, these conditions need be reliable only in *this* world and worlds nomologically like this one. Psychology is beholden to worlds with the psychological (and related) laws of *our* world; other nomologies need not concern it. In sum: for a property theory to be in the offing, there must be *something* that keeps computation and content in phase, such that computational states/processes track intentional states/processes most of the time—but the tie need not be perfect nor hold in all possible worlds. These conditions can be met far short of sufficient conditions as conceived in classical conceptual analysis.

Unfortunately, the revised Premise One immediately suggests a revised Argument for Incompatibility. Frege cases (such as 'the Morning Star' and 'the Evening Star', 'Mark Twain' and 'Samuel Clemens') appear to be examples of computational type distinctions to which no content distinction corresponds; Putnam's Twin Earth and Expert ('elm'/'beech') cases appear to be examples of content type distinctions to which no computational distinction corresponds. (Fodor also discusses what he calls 'Quine cases': 'rabbit' and 'undetached rabbit part', for example. We'll introduce Quine cases later.) Because of Frege, Twin Earth, Expert and other cases, it is tempting to think that:

> *Premise Three (Revised)*: If content is relational then the coinstantiation of the computational-implementer and intentional implemented will not be reliable.

Premise Two and Premise Three (Revised) entail that coinstantiation *will not be* reliable. Infointentionalism, computationalism and Premise One (Revised) together entail that coinstantiation *must be* reliable. The ancillary premises seem solid. So it appears, once again, that infointentionalism and computationalism are not consistent.

In response to this variant of the Argument for Incompatibility, Fodor goes after Premise Three (Revised). Roughly speaking, Fodor argues that— Frege, Twin, Expert and other cases notwithstanding—computational

states/processes and intentional states/processes *do* reliably coinstantiate, at least when it matters. So the revised Argument for Incompatibility is also unsound. We turn, at last, to Fodor's defence of this claim.

4. Complications: Tying Content to Computation

When content is construed informationally—that is to say, as a matter of syntactic forms being in relationships to external objects—there are two broad ways in which content and computation could come apart, Fodor suggests.

(1) There could be computational distinctions that do not reflect differences of informational content (Frege cases), and,

(2) There could be differences of informational content that are not reflected in difference of computational state.

The latter in turn might happen in two ways:

(2a) The difference of content is not available to the cognitive system, as in Twin Earth (and also Expert) cases. (In Expert cases it is available to an expert: Fodor cannot tell a beech from an elm but an expert can. In Twin Earth cases, it is not available to anyone.)

(2b) The difference is available to the cognitive system, but it cannot be captured in a purely informational theory, as in the case of Quine's rabbit/undetached rabbit part.

For each of these ways in which disconnection might seem possible, (1), (2a) and (2b), Fodor tells us that there is a mechanism tying content and computation together. To explain what he has in mind, he offers an analogy. Why are appearing to be a dollar bill and actually being a dollar bill tied together—not perfectly but very, very reliably? Because of the mechanism of the police stamping out counterfeiting.[2] We would then expect Fodor to say, 'And here are the mechanisms for content and computation for each of the three kinds of case.' But that is not what he does at all. Instead, the analysis goes off in a curious direction, indeed in two curious directions:

(i)   Instead of identifying three policelike mechanisms, Fodor offers us in effect *a series of explanations of why we do not need one*, at least in two of the three kinds of case. (The problem in his treatment of the third is even bigger, as we will see.)

---

2   Fodor insists that, like the police, the mechanism tying content to computation must be at work currently; it is synchronic with what it ties together. Thus evolution can be no part of the story. We will see another example of his lack of enthusiasm for evolutionary accounts later.

The second curiosity arises from the fact that Fodor uses at least two variations on the content-as-information theme, which he calls causal and nomic. (We'll untangle these terms shortly.) But:

(ii)   The explanations mentioned in (i) go through straightforwardly only for the causal variant.

4.1 Mechanism or Explanation?

The first curiosity is: instead of mechanisms, in two of the three cases Fodor offer us explanations of why we do not need any. These are the Frege and Twin Earth cases. (As we said, the problem with his treatment of Quine cases is different.) Let us begin with Frege cases. Take Oedipus and his unfortunate affair with his mother. Fodor says that, so far as *content* is concerned, the proposition thought by Oedipus to be true—that he was marrying Jocasta—and the proposition thought by Oedipus to be false—that he was (gasp!) marrying his mother—are the same: 'mother of Oedipus' and 'Jocasta' have the same reference so carry the same information. Nevertheless, they're computationally distinct and clearly had or would have had different effects on Oedipus.[3] If so, content doesn't map one-to-one with computational state.

But cases like Oedipus are not a problem, says Fodor, because they are *unsystematic accidents*; if it happened regularly that we did not know that coreferring terms referred to the same object, practical reasoning would become useless. Indeed, it takes such complicated circumstances to make a story like Oedipus's plausible that such cases would have to be rare. If these cases are rare accidents, however, all they show is that content and computational state are not perfectly linked. They do not show that the two are not *reliably* linked, and this is all we need. So far so good. But now ask: where's the mechanism? If accidents are allowed, we do not need to block them and so we need no content/computation mechanism to block them; there would be nothing for such a mechanism to do. Curious.

'Surely', it will be protested, 'you have missed Fodor's point. We *also* need something to explain how it is that Oedipus cases *are* unsystematic—that we do generally know that our coreferring terms refer to the same object. *This* is the mechanism Fodor has in mind.' That this is what Fodor has in mind is, at the very least, not obvious. Let us follow his analysis. He starts by saying that:

> 'Intentional systems' invariably incorporate mechanisms which insure that they generally know the facts upon which the success of their behaviour depends. (Fodor, 1994, p. 48)

---

[3]   The two states are distinct because they are encoded in two syntactically distinct sentences of Mentalese.

Is this the mechanism we are looking for? No; for 'rational agents' to 'reliably make a point of knowing the facts that the success of their behaviour depends upon' (Fodor, 1994, p. 46), the facts have to cooperate. We have to have what poor Oedipus lacked, information adequate to know the truth of the relevant identity statements. As it happens, we generally do, because 'the syntactic structure of a mode of presentation reliably carries information about its causal history' (Fodor, 1994, p. 54). And do they also carry information about other syntactic structures, that they refer to the same object?, we might ask. But that is not the vital question. The vital question is this: Does any *mechanism* guarantee this reliability? Indeed, does anything *guarantee* it? Fodor says nothing to suggest that it is more than a happy accident, though one essential to practical reasoning, indeed probably to life itself.

If there is a mechanism producing the happy covariations that are supposed to allow us to know, most of the time, the identity of reference of our coreferring expressions, the only candidate that we can think of is natural selection. In some contexts, Fodor expresses a distinct lack of enthusiasm for evolutionary arguments (1994, p. 20 for example), but this may be one place where they can do some work. Whatever, it would be peculiar to call anything that could be involved here a *mechanism*, certainly if the police arresting counterfeiters is an example of a mechanism. Compare: 'Why aren't there lots of elephants on Lake Ontario today? Well, you see, there's a mechanism . . .'.

We discover the same pattern even more clearly in Fodor's treatment of Twin Earth cases. Fodor's response to them is to claim that they do not occur in our world and will not occur in any nomically near one (Fodor, 1994, pp. 38–9). He may well be right; but if he is, we do not need any *mechanism* to deal with them. Once again, instead of identifying a mechanism, Fodor has shown that we do not need one.

That leaves the Quine cases. (We will not consider Expert cases.) They need a word of explanation. The problem that 'rabbit'/'undetached rabbit part' ('urp') cases pose for informational semantics Fodor-style is important. The terms 'rabbit' and 'urp' clearly have different contents. Indeed, since a rabbit is more than an urp, if one term correctly applies, then the other does not. Yet they are always coinstantiated. Given that, the *information* contained in 'rabbit' and 'urp' is the same, on Fodor's stringent notion of information. From this it follows that no semantics based purely on such a notion of information is going to capture the difference of content between the two terms. Fodor's suggestion is that we can capture the difference between 'rabbit' and 'urp' if we add a notion of 'inference potential' (shades of conceptual role!). In particular, by checking whether an Informant (Inf) will accept or reject certain conjunction reductions, we can tell whether she uses 'rabbit' to mean *rabbit* or *urp*, 'urp' to mean *urp* or *rabbit*.

Here is how the story goes. (Fodor tells it in terms of triangles and triangle parts but we will tell it in its rabbit and urp version). In addition to a rabbit and an urp, consider also, say, the front half of a rabbit. An appropriately located urp could be part of both a rabbit and the front half of a rabbit (an

undetached ear, eye, or nose would be some examples). Now we ask, does 'rabbit' mean *rabbit* or *urp*? Does 'urp' mean *an undetached rabbit part* or *rabbit*? Being the front half of a rabbit excludes being a rabbit but being an (appropriate) urp is compatible with being part of both a rabbit and the front half of a rabbit. Thus, if 'rabbit' and 'front half of a rabbit' mean *rabbit* and *front half of a rabbit*, Inf will never, it seems, accept 'A is a rabbit and a front half of a rabbit'. On the other hand, if 'rabbit' and 'front half of a rabbit' mean *rabbit part* and *part of the front half of a rabbit*, Inf will sometimes accept 'A is a rabbit and a front half of a rabbit'. So all we have to do is check and see which predicate conjunctions Inf accepts and we can determine what she means by 'rabbit' and 'urp' (1994, p. 73). So far so good.

But now a problem appears—a problem that arises from the very nature of conceptual role semantics. There's an intuitive content difference between 'urp' and 'rabbit'. Fodor needs to capture this difference. But what he suggests as the difference *in content* is, we fear, just a syntactic difference. To see this, ask yourself what exactly the content difference is meant to be. We repeat: it cannot be anything informational because Inf's information about rabbits and about urps is the same; 'rabbit' and 'urp' covary perfectly with both rabbits and urps. Recognizing this, Fodor proposes that the content difference is a matter of inference potential. We grant that there is a difference in inference potential between 'rabbit' and 'undetached rabbit part': 'x is a rabbit' entails certain things that 'x is an urp' does not. But there's also a difference in inference potential between 'rabbit' and 'instance of rabbithood'—and these differences in inference potential are not the same. To take an example, call the difference between 'rabbit' and 'urp' InfPot_1. Now consider the difference in inference potential between 'rabbit' and 'instance of rabbithood'—which we'll call InfPot_2. Two questions arise. First, what exactly is InfPot_1? How can we characterize it in such a way that it's appropriately different from every other inferential potential, including InfPot_2? The foggy nature of conceptual role semantics makes this a terrifically difficult question to answer. Second question: what mechanism ties this content—i.e. InfPot_1—to 'urp'? The painful answer is: it's entirely mysterious how InfPot_1 gets tied to 'urp', rather than being tied to 'instance of rabbithood'. In sum, we can't help thinking that the only difference that Fodor can point to between 'undetached rabbit part' and 'instance of rabbithood' is something syntactic. But nothing purely syntactic adequately captures the intuitive content difference. If this is right, then Fodor is making syntax serve as *part* of content. Moreover, so far as we can see, nothing ensures that this syntactic difference corresponds to, or will stay in synch with, the intuitive content difference. The Quine cases remain, alive and well.

Does Fodor have any escape? If conjunction reduction proclivities are to be part of content, they must at least covary with the relevant differences of content. We can think of two possibilities:

(1) Fodor could simply *define* incompatibility of satisfaction as an Inf's unwillingness to conjunction reduce.

This seems a desperate expedient: what does the fact that 'rabbit' is satisfied only when 'urp' is not have to do with some rule for transforming uninterpreted symbols buried deep in the brain? But out of desperation grows an idea. What if,

(2)   we could find a *mechanism* that tied the computational reduction to the semantic difference? This should be the escape Fodor wants, given the rest of *E&E*. Can we think of one?

It does not look promising. Strangely enough, Fodor is no help here; indeed, he does not even mention a mechanism in connection with Quine cases! This gap and difficulty of bridging it may be an instance of a wider problem. At the beginning of *E&E*, Fodor tells us that if computations are to house content, computations must track truth in the way that our reasonings do. To see the bigger problem first, suppose that Fodor is right about everything and there is no way in which symbol structures and contents might systematically come apart, not in this and near worlds at any rate. Would that be enough to ensure that computation will *track truth* reliably? It is not clear that it would; even if computation and content *nodes* line up correctly, the *relationships among them* could still come apart.[4]

Now apply this bigger problem to conjunction reduction and the problem of covarying terms that are not jointly satisfiable. The problem here is to find something to ensure that a purely computational move tracks mutual nonsatisfiability correctly. It's not the same problem, but it is a closely related one. And Fodor says as much about it, namely, nothing.

This problem is far more serious than the problems we identified earlier in connection with Frege and Twin Earth examples. There Fodor promised us mechanisms and instead gave us arguments for why we don't need them—a fairly parochial failing. Here he produces no mechanism *and* no argument that we do not need one. This looks more serious.

But what looks like a serious setback—i.e. the inadequacy of inference potential as an answer to Quine—may point to a better solution. Here's why. Whether the inference potential of 'urp' tracks the intuitive content difference between it and 'rabbit' or not, the introduction of 'inference potential' as an element of content raises troubling issues on its own. Specifically, though Fodor argues that this move invites only a benign form of holism, it's surely very hard to help oneself to just a soupçon of conceptual role semantics. (Fodor says that he can isolate the 'logical syntax' of conjunction reduction from the rest of language, but his treatment of this seems a little blasé. In particular, to make the separation, Fodor would have to be able to

---

[4]   It is interesting that, after raising the problem of tracking truth in Chapter 1, FOdor never returns to it. This is another place where an evolutionary account might do some work.

separate conjunction reductions based on syntax from reductions based on evidence, i.e. purely on information. Perhaps he can; but given Quine's worries about the very possibility of doing such things, we'd like to see the argument.) Given the problems with borrowing from conceptual roles, it would thus be better to go entirely informational. And we can find a way to do so—if we allow information to be more than mere covariance.

Here is how. The relationships of a rabbit, or being a rabbit, to our syntactic structures and of an urp, or being an urp, to our syntactic structures differ in more ways than just covariance, ways just as naturalistic. In particular, many of the *causal* relationships are different, as different as the very different causal powers that go with being a rabbit and being an urp. Fodor urges that whenever there is a law connecting being a rabbit and tokening 'rabbit', there will be a law connecting being an urp and tokening 'rabbit'. At the appropriate coarseness of grain, he is right. However, many of the causal details of the connections will differ. These differences can ground differences of information just as well as covariance can. And where covariance cannot do the job, as with 'rabbit'/'urp' and other Quine cases, these other differences can step in. That removes the problem caused by Quine cases.

Fodor does not avail himself of this solution only because his notion of information is excessively stringent. For him, the *only* way in which the information carried by two terms can vary is if there are contexts in which the terms do not covary; and the only way they can do that is if they refer to objects that do not always covary. But content is also determined by what *undergirds* the covariance: 'dog' has the content it has not simply because it covaries with dogs, but also because tokens of it are *caused* by dogs (see Quine, 1960, p. 30). This is entirely compatible with a naturalistic, externalist account of content. Assuming—as seems plausible—that it is the causal powers of *rabbits* that maintain the covariance between 'rabbit' and rabbity stuff and that the causal powers of urps are relevantly different, 'rabbit' will mean *rabbit*, not *urp*. In which case, 'rabbit' meaning *urp* is excluded without appeal to conjunction reduction or any other kind of 'inference potential'. And the Quine cases would no longer be a problem. True enough, content so construed will go beyond covariance, but we cannot imagine why this should concern Fodor.

4.2 *Variations on the Notion of Information*

Above we urged that, just when Fodor seems poised to describe the police-like 'mechanism' that keeps content and computation reliably in phase, his discussion goes off in peculiar directions. Having explored one of them— no mechanism, just explanation, and in Quine cases no explanation either—, we turn now to the other one. Fodor acknowledges two variations on the content-as-information theme: causal and nomic. Indeed, he explicitly discusses the difference between these, in Appendix B. Here's what's strange: the explanations Fodor gives in connection with Frege and Twin Earth cases go through straightforwardly only for the first.

Fodor calls the two conceptions the *causal* informational and the *nomic* informational. Fodor calls the first the historical conception (1994, pp. 115–9) but by 'historical' he means something quite different from the biographical/historical conception propounded by Dretske (1993) and others. Dretske argues that AI systems do not and could not have content, intentionality, etc. simply because they have the wrong kind of history. On this view, even if a system were to be built, to the appropriate fineness of grain, exactly like us in all relevant respects, and even if it behaved exactly like us, the difference in its history would ensure that it does not have content—even though we do. The only history Fodor has in mind, however, is the causal history of a word token, i.e. the link between it and the thing that caused it to occur.

Fodor calls the causal informational notion of content the view that, 'the content of mental representations is constituted by their etiology' (Fodor, 1994, p. 82). As we've seen, Fodor goes to great lengths to show that content thus construed is compatible with computationalism. However, there is also the nomic notion. On the causal story, a computational state comes to carry information by entering a causal relationship with some object: a token of 'dog' comes to carry the information contained in the concept DOG by being in a causal relationship with a dog. On the *nomic* informational story, in contrast, a computational state comes to carry information by satisfying certain counterfactuals: a token of 'dog' as found in me carried the information *dog*, for example, if I *would* say or otherwise token 'dog' *were I* to be in the presence of a dog. It is not necessary that my token of 'dog' was ever *actually* in a causal relationship with a dog.

As Fodor brings out very nicely in Appendix B, these notions of broad content are very different from one another. Fodor centres his discussion on Davidson's (1987) old friend the Swampman, the molecule for molecule duplicate of Davidson who springs to life ('life'?) one day in a swamp. Intuitively, Swampman seems to have content at the very start of his new life, i.e. prior to entering into causal relationships with objects. Now, this is also what the nomic story would entail, and it is story Fodor embraces. However, on the *causal* story, we would have to deny that the Swampman had content—as a number of theorists have done. So the two conceptions are quite different. So far so good.[5,6]

---

[5]   At this point someone might object: 'the dispositional, nomic account runs into as much trouble over Swampman as the causal one. Consider worlds in which the counter-factuals—the ones that specify the dispositions in question—do not hold. In these worlds, the Swampman would not have content. So the nomic account runs into much the same trouble with the Swampman. They are not as different from one another as you are making out'. Compare: glass would not be brittle in worlds in which moving masses do not carry kinetic energy, and tokens of 'dog' would not have the content *dog* in worlds in which dogs do not cause tokenings of 'dog'. Response: who knows what the intentional psychology of creatures in a world nomically so different from ours would look like? All that matters is that the dispositional account work in our world, and worlds nomically near. And in these worlds, the two notions yield dramatically different pictures of semantic content.

Now ask a question that Fodor does not ask. As we've seen, Fodor argues that—Twin, Frege and Quine cases notwithstanding—the link between content and computational implementation is reliable. Fodor bases this story on the causal notion of content. How would it go on the nomic notion? Our answer is: quite differently, for some of these cases at least; and the differences do not work to Fodor's advantage. He gives no indication that he sees either point.

Start with the Frege cases. However well Fodor's story about why they are no problem works on the causal informational account, it does not work at all on the nomic informational version. Here is why. If content is a matter of actual causal connection to an object, then it is perfectly possible to locate what is common to Oedipus's two beliefs—it consists in their being linked causally to the same object and this becomes the content of both 'Jocasta' and 'Mom'. Because the two words are syntactically distinct, we have a case of computation/content disconnection, but that's okay because it's accidental. Switch to the nomic notion of content as dispositions of some kind, however, and this story collapses. There now has to be some single disposition in Oedipus that gives content and the same content to 'Jocasta' and to 'Mom'. What could it be? Certainly not a disposition to get married! Nor, to be more serious, the respective word tokening dispositions either. Until Fodor identifies such a disposition, he has no story about Frege cases on the nomic account of information.

What about Twin Earth cases? Again a problem. Where information derives from *aetiology*, the apparent difference in content between (say) Adam and Twadam is that Adam's tokens of 'water' are causally linked to H_2O, Twadam's to XYZ—a substance phenomenally indistinguishable but chemically (or something) distinct from H_2O. Here there is a clear sense in which the two tokens of 'water' have different contents, even though elicited by indistinguishable environments: one is causally linked to H_2O, the other to XYZ. Different contents, same computational state. (And, says Fodor, it doesn't matter because no Twin Earth is 'nearby'.)

However, things again turn sour when we go nomic. On the nomic story, the difference in content between Adam and Twadam is presumably that Adam would token 'water' in the presence of H_2O while Twadam would token 'water' in the presence of XYZ. However, Adam's disposition would also lead him to token 'water' in the presence of XYZ and Twadam's disposition would lead *him* to token 'water' in the presence of H_2O. Thus there is *no difference* between their two dispositions. On the causal story, recall, the difference of content is the result of an actual link, the causal link either

---

6   Rather than saying that we have different *notions* of content, another approach would be to say that we have different *theories*, at least three of them, of where content, conceived of in a single conception, derives from. The latter approach would require that the causal, nomic and historical notions of content have something both relevant and substantive in common. Fortunately, we do not need to settle this question. Three conceptions or three theories, the consequence for Fodor are the same.

---

to H_2O or XYZ. Absent that link and there is nothing that could make the two in any way different, not so far as content is concerned. If so, there is no difference of content between Adam and Twadam on the nomic theory in the first place. Hence there is no room for any same-computational-state-different-content worry. On the nomic notion, that is to say, Twin Earth cases are not even an apparent problem for Fodor.

Objection to our objection: 'Construe dispositions *broadly* and there *is* a single disposition which gives the same content to 'Mom' and 'Jocasta'; likewise, there *are* two distinct dispositions for the single term 'water', here and on Twin Earth. Oedipus has a disposition *with respect to the person Jocasta*, which gives the same content to both 'Jocasta' and 'Mom'; and Twadam has a disposition to token 'water' linked to XYZ; while Adam here on Earth has a disposition to token 'water' that is importantly different because *it* is linked to H_2O. That is to say, there is a single disposition for the single content of 'Jocasta' and 'Mom' in Oedipus and distinct dispositions (one XYZ-linked, the other H_2O-linked) for the distinct contents of the single mentalese expression 'water' in Adam and Twadam. Moreover this appeal to broad dispositions would be entirely in the spirit of Fodor's new approach. Actual observation of a speaker won't make clear which disposition the speaker has; but only an unreconstructed behaviourist would worry about that.

We have two responses:

(1) *What* single disposition does Oedipus have with respect to Jocasta? The only disposition(s) we've heard mentioned is/are the disposition(s) sometimes to token 'Jocasta', sometimes to token 'Mom'. The two could be unified with respect to content only by being linked by *actual lines of causality* to a single woman—at which point we have the causal picture again and no longer a nomic one. Similarly, how could Adam's disposition be in any way *different* from Twadam's— prior to tokenings of 'water' actually being caused in the two of them by the two different substances?

(2) Next ask yourself: Would any even *counterfactual* observation unite Oedipus's disposition(s) or distinguish between the XYZ and the H_2O-linked dispositions in Adam and Twadam? We can think of no either actually *or counterfactually* observable behaviour that could do either job. It seems to us, however, that it is part of the nature of dispositions that they enter into what an agent *would be observed to* do: a 'disposition' which makes no difference either to actually or to counterfactually observable behaviour isn't really a disposition at all.

In short, we do not think that there can be broad dispositions in the sense required. Hence appeal to broad dispositions cannot bring together the causal and the nomic stories; our concern that Fodor's story does not work for the nomic notion of content stands.

To sum up this section: On the nomic view, Twin Earth cases are actually easier for Fodor than the causal one, though for reasons that he might not

relish. If he wants his story of Frege cases to work for the nomic view as well as it does for the causal one, however, he needs a rather remarkable kind of disposition, of which he owes us an account. Curiously, his treatment of Quine cases seems to work as well, or as badly, on either view.

5. *An Epilogue on Motivation*

So much for curiosities and complications. Despite their power to sway us, Fodor might very well be right that Frege, Twin Earth, and Quine cases pose no threat to reconciling an informational picture of content with a computational account of cognition—though the claims about mechanisms, etc. that we have been examining leave us puzzled. One might stop here. But there is a huge issue beneath the details of Fodor's account that we want to examine briefly. What *motivates* Fodor and others to work so very hard to reconcile content-as-information and computationalism? We will end with some remarks on this question.

What motivates Fodor to show that intentionalism and computationalism are compatible is naturalism: he wants to naturalize the intentional, i.e. provide a naturalistic, nonintentional implementation of intentional states. Naturalism, Fodor notes, is also a prime motivation for embracing information theoretic semantics. Probably it is also a main motive for computationalism: when Fodor says that computation is the only 'remotely serious' notion of process capable of implementing truth preserving state transitions, he probably means 'plausible and *naturalistic*'. At bottom, then, what Fodor describes is a conflict between intentionalism and the only naturalistic account of human intentionality that he knows of—one in terms of information and computations. But if computationalism plus information theoretic semantics is the only way to naturalize intentionality, then infointentionalism and computationalism had better be compatible. If they are not, that would be a *very bad thing*. The reason is, chez Fodor, if we cannot naturalize intentionality, we cannot have it:

> Psychologists have no right to assume that there *are* intentional states unless they can provide, or anyhow foresee providing, or anyhow foresee no principled reason why someone couldn't provide, naturalistic sufficient conditions for something to be *in* an intentional state. (Fodor, 1994, p. 5)
> ... if the problems about implementation we've been discussing are real and not solvable, only the elimination of the intentional would be a cure adequate to the disease. (Fodor, 1994, p. 15)

If we cannot show that content is natural, then we must give up intentionalism. If the intentional is eliminated, then evidently there aren't any intentional causes, intentional processes or intentional laws. And, Fodor says, 'if there are no intentional laws then there are no psychological expla-

nations' (Fodor, 1994, p. 3). This is why Fodor goes as far as to say that failure to naturalize content would lead to 'the greatest intellectual catastrophe in the history of our species' (Fodor, 1987: xii).

It is this last step—that failure to naturalize must lead to elimination—that we find less than obvious. In contrast to Fodor, there is clearly something in the view—now familiar from work by Stich (1992), Tye (1992), Horgan (1994), Stich and Laurence (1994), Rudder Baker (1995) and others—that nothing earth shattering need happen if we cannot give a naturalistic account of intentionality. In particular, our inability to achieve this goal would not entail that intentional states and processes are not real. The following is surely possible: intentional states and processes play a sufficiently central role in ordinary life and scientific theorizing that we must be ontologically committed to them—whether we can naturalize them or not: i.e. whether we can specify, in nonintentional terms, the conditions for being in an intentional state.[7] Given this possibility, in so far as intentional states *are* central (and, in our opinion, they're as central as could be) they are safe from elimination. Hence fear of eliminative materialism should not be taken to motivate the reconciliation of information theoretic semantics with computationalism.

Indeed, the pressure to naturalize can even get in the way. Stich et al. are quite happy to study intentional, semantic, and other phenomena on their own terms and in all their richness. Fodor, by contrast, always has one eye (at least one eye!) on what a naturalistic account of these phenomena would look like. That sometimes leads him to oversimplify the phenomena and gloss over complexities. Perhaps he's paying too much attention to the phantoms of San Diego.

In sum: Fodor's strategy in *E&E* is to urge that psychological states and processes can be both ineliminably intentional (information bearing) and implemented computationally—without there being computationally sufficient conditions for the instantiation of intentional properties. All we need is that computational states/processes reliably coinstantiate with intentional states/processes in our world and worlds nomically near to ours. Which, despite the problems with Fodor's arguments, they may do. But Fodor argues for the thesis in some some strange ways and owes us some additional arguments. We also wonder about Fodor's motivation for taking on the project. Do we need to show that cognition-as-computation and con-

---

[7] This might happen in any number of ways. Here are two. The relationship between content and computation might be so complex as to be unfathomable by humans. If that were the case, naturalizing would be 'possible in principle' only in a most attenuated sense of those words. Alternatively, and more radically, the 'constitutive principles' in the two domains might be fantastically different—though each is scientifically tractable in its own terms. (Anyone recall anomalous monism?)

tent-as-information fit well together to continue to believe in the intentional? Probably not.

Department of Philosophy and Institute of Interdisciplinary Studies
Carleton University

References

Davidson, D. 1987: Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, 61, 441–58.
Dretske, F. 1993: The Possibility of Artificial Intelligence. Paper presented to the American Philosophical Association, Eastern Division, 28 December 1993.
Fodor, J.A. 1987: *Psychosemantics*. Cambridge, MA.: MIT Press.
Fodor, J.A. 1994: *The Elm and the Expert*. Cambridge, MA.: MIT Press.
Horgan, T. 1994: Computation and Mental Representation. In S. Stich and T. Warfield (eds) *Mental Representation: A Reader*. Oxford: Blackwell, pp. 302–11.
Quine, W.V.O. 1960: *Word and Object*. Cambridge, MA.: MIT Press.
Rudder Baker, L. 1995: *Explaining Attitudes: A Practical Approach to the Mind*. Cambridge University Press.
Stich, S. 1992: What Is a Theory of Mental Representation? *Mind*, 101. Reprinted in S. Stich and T. Warfield (eds) *Mental Representation: A Reader*. Oxford: Blackwell, 1994, pp. 347–64.
Stich, S. and Laurence, S. 1994: Intentionality and Naturalism. *Midwest Studies in Philosophy*, 19.
Tye, M. 1992: Naturalism and the Mental. *Mind*, 101, 421–41.