

The Maribor Papers in Naturalized Semantics, edited by
Dunja Jutronic, Maribor University Press, Slovenia, 1997, pp. 157-74.
[Proceedings of the *Symposium on the Naturalized Semantics and its Methodology*,
Maribor, Slovenia, June 10–15, 1996.]

Pure Informational Semantics and the Narrow/Broad Dichotomy

MURAT AYDEDE

The University of Chicago
Department of Philosophy
1010 East 59th Street
Chicago, IL 60637, USA

(m-aydede@uchicago.edu)

ABSTRACT

The influence of historical-causal theories of reference developed in the late sixties and early seventies by Donnellan, Kripke, Putnam and Devitt has been so strong that any semantic theory that has the consequence of assigning disjunctive representational content to the mental states of twins (e.g. [H₂O or XYZ]) has been thereby taken to refute itself. Similarly, despite the strength of pre-theoretical intuitions that exact physical replicas like Davidson's Swampman have representational mental states, people have routinely denied that they have any intentional/representational states. I want to focus on a particular brand of causal theory that is not historical, the so-called pure informational or nomic covariance theories, and examine how they propose to handle twin cases and replicas like Swampman. In particular, I will take up Fodor's version of the theory, since it is the best worked out specimen in this genre. I will argue that such (non-historical/non-teleological) theories as Fodor's are bound to assign disjunctive content to twins and representational content to replicas. I will also argue that this consequence should perhaps be welcome. I will end by sketching a picture according to which a pure informational semantics can accommodate both the internalist and the externalist intuitions.

1. Introduction

After Putnam’s celebrated thought experiment involving Twin Earth, the role of actual context or environment of intentional agents has been taken to be essential in fixing the reference or truth-conditions of those agents’ contentful mental states, that is, in fixing what it is that those states are about or represent. Indeed, the intuition that twins’ mental states have different truth-conditions, thus different (broad) contents, is alleged to be so strong that any naturalistic theory of reference, or as I shall put it, any theory of broad content, that assigns the same truth-conditions to twins (whatever they may be) is thereby taken to refute itself. Almost everyone in the field takes this result to be a datum to be explained or accommodated by any theory of broad content.

Historical-causal theories of reference, developed by Kripke, Donnellan, Putnam, Devitt and others, have an explanation for the source of these intuitions: namely, the reference of many terms like proper names and natural kind terms is fixed by some sort of an actual causal interaction with the referent and the user of the term. It is in virtue of our actual causal interactions with the local stuff that our water-thoughts are said to be about H₂O, but not about XYZ.

It is almost universally held that it is not just historical-causal theories that can account for these intuitions, but any causal theory that attempts to explain in naturalistic terms what it is that makes our thoughts about things in the world must respect these intuitions. Devitt in his book (1996) divides “naturalistic attempts to explain the direct noninferential links to reality” into three groups according to the kind or nature of causal relation hypothesized to underwrite the relations between representations and reality.

Naturalistic attempts to explain the direct noninferential links to reality that determine the reference of a token have appealed to one or more of three causal relations between representations and reality: (i) the *historical* cause of that particular token (the idea for this arises from Kripke 1980, Donnellan 1972, and Putnam 1975); (ii) the *reliable* cause of tokens of that type (the idea for this was first suggested in Stampe (1979 and

Dretske 1981); and (iii) the purpose of tokens of that type, or the mechanism that produces them, where the purpose is explained causally along Darwinian lines — the *teleological* cause (the idea for this was first suggested in Papineau 1984 and 1987 and Millikan 1984). (Devitt 1996, p.161)

Devitt's terminology here is a little non-standard. I will call the latter two varieties species of informational semantics and focus only on the purely informational one. By this I mean to exclude the Millikan/Papineau type of approach where the information-carrying function of representations is shaped by the actual history of the causal interaction between the symbol and its referent. When put this way, Dretske's recent proposals are closer to Millikan's than to Fodor who is running a version of what I'll call Pure Informational Semantics (PIS). In what follows, I'll focus on Fodor's pure informational semantics, but only as an example to the kind of approach I have in mind, where it is not the actual historical etiology of a symbol that fixes its referent, but rather the currently existing nomic correlations. As such, informational approaches seem to stand in stark contrast with historical-causal theories of reference.

As I said, whatever the nature of causal relations are, whether actual-historical or just merely nomic, the unanimous consensus is that all such theories do in fact assign, or at any rate, ought to assign different broad content to twins' relevant mental states. In what follows, I will argue that it is not at all clear that pure informational semantics does so. I will also raise and discuss towards the end the issue of whether it ought to do so.

Here is how I propose to proceed. I'll first take up a specific version of pure informational semantics, namely that of Fodor, and show that it doesn't succeed in assigning different broad contents to twins' mental states. In particular I'll show that it assigns to them disjunctive content, H_2O or XYZ . The reason I have chosen Fodor's version is because it is the most fully worked out theory that is, in certain ways, especially designed to handle the disjunction problem. But nothing hangs on this particular choice. After I finish with Fodor I'll make some general points about the bare essentials of what a purely informational semantics is supposed to be. I'll discuss how to interpret the informational accounts in connection to

twins' mental states. We'll see that there are problems about how to interpret the informational contents of twins' mental states. At that point, I'll enumerate some of the options we might adopt in dealing with the informational contents of twins' states. I'll point out that at the hands of Fodor, the notion of broad content that his PIS delivers turns out to be extensionally equivalent to a notion of narrow content conceived as a mapping from contexts to broad contents, a notion that Fodor had been defending vigorously until quite recently.

I will argue that this situation is in fact quite suggestive in certain ways. I'll end by painting in very broad brushes a picture about the relationship between PIS and broad/narrow dichotomy, and suggest that a pure informational semantics has perhaps the resources to adequately explain both internalist and externalist intuitions in psychosemantics. Let me then start with Fodor's version of PIS.

2. Fodor's Pure Informational Semantics

According to Fodor's Pure Informational Semantics (PIS), a Mentalese predicate #F# expresses the property F IF

1. #F#s nomically covary with Fs,
2. for all $G (F)$, if #F#s nomically covary with G_s qua G_s then #F#'s nomic covariation with G_s is (synchronically) *asymmetrically dependent* on #F#'s nomic covariation with Fs.
3. #F# is actually caused to be tokened by non-Fs (i.e. tokenings of #F# are robust. (Cf. Fodor, 1990a:117–9, 121)

Since the content-making correlations are stated purely counterfactually, there is a clear sense in which the theory *prima facie* assigns broad content to symbols *independently of any particular contexts, or equivalently, without regard to what particular context the subject happens to be actually in*. The theory *prima facie* assigns broad content to a symbol-and-mechanism pair on the basis of the counterfactual behavior of that pair, i.e. on the basis of what it would nomically covary with across different contexts. I'll argue in a moment that what is *prima facie*

the case remains so even after Fodor's attempts to block this consequence of his theory. However, let me focus, for the moment, on the theory itself.

According to the theory, if you want to know what the broad content of a mental state is, it is not enough to look at what it actually covaries with in its present context: you also have to look at its counterfactual behavior.¹ So to individuate a mental state on the basis of its broad content is *prima facie* to individuate a certain set of internal dispositions supervenient on the brain that would cause the mental state to covary with a range of external stuffs in different contexts. But, as we'll see below, this turns out to be exactly the characterization of narrow content given by Fodor.

Furthermore, what makes the counterfactuals of this purely informational theory of broad content come true is (mostly)² in the head, and importantly so: it is certain mechanisms that fix the disposition to token a certain symbol when a certain property is suitably instantiated.³ Put differently, instantiation of a certain property nomically controls the tokening of the symbol that expresses the property. But since this controlling is nomic, i.e. counterfactually fixed, there is no requirement that the property has ever actually been instantiated. It is rather: if it *were* instantiated it *would* cause the tokening of the symbol that expresses it.⁴

¹ Cf. Fodor's discussion of how God would know what the broad content of one's mental state is (1990a:125–7).

² This qualification may be thought tendentious. But what I have in mind here is Fodor's attempt to summarily extend his purely informational theory to cover Mentalese proper names like #Aristotle# and deferential concepts like #elm# and #beech# (e.g., 1994:118–9; cf. also 1991:285–6) for which extra-cranial social/linguistic mechanisms (like deferring to experts) are viewed as mechanisms that sustain the counterfactuals. Fodor's remarks are rather very sketchy and restricted to a few scattered brief passages in his (1994). I think this extension of the theory is ultimately untenable. But I'll have to leave its discussion for some other occasion.

³ Fodor is explicit about the intra-cranial dispositional character of his theory: "At a minimum, an informational semantics ... takes the content of one's concepts to be constituted by one's dispositions to apply them" (1994:30). Here is another such quotation (among others): "What your words(/thoughts) mean is dependent entirely on your *dispositions* to token them (on what I called the "subjunctive history" of their tokenings), *the actual history* of their tokenings being semantically irrelevant" (1990a:120).

⁴ For the most part, I will be assuming in what follows that all the asymmetric dependences are in place (i.e., that the second clause of Fodor's informational semantics is satisfied) in order to focus on primary content-making nomic relations.

It is the truth of such counterfactuals that fix the broad semantics of mental symbols. Indeed, this is how Fodor wants to account for the semantics of concepts like UNICORN and CENTAUR.⁵ No unicorn or centaur has ever actually existed, and it is likely that none will actually exist. But this is irrelevant. As long as they are at least nomologically possible, there are worlds in which my #unicorn# token is actually controlled by unicorns. It is this counterfactual truth, according to Fodor, that makes my #unicorn# symbol mean *unicorn* in *this* world, i.e. in a context in which there are no unicorns. In fact, the sense in which there are no unicorns is stronger than the sense in which there are no, say, pink submarines. We can easily instantiate the latter property, but I don't think the former can be instantiated in the same sense. There are many nomologically possible properties that won't ever be instantiated in this world. Fodor wants to use the same counterfactual apparatus to explain how we can have concepts that express them, even though our actual world is a context in which there are (and, in a loose sense, can be) no instantiation of these properties.

Let me also point out one important aspect of Fodor's version of informational semantics. If, as Fodor says, there is nothing to the meaning of a symbol except its denotation or truth-conditions, then the pure informational account of broad content in terms of what internal dispositions sustain which covariations is in a non-trivial sense an internalist naturalism, contrary to what Fodor himself says in his (1994).⁶ In other words, what Fodor seems to be offering is an internalist theory of broad content exclusively on the basis of nomic relations that obtain between the world and the mind. I know this sounds paradoxical,

⁵ See Fodor (1987:163–4:fn.5) and (1994:115–19).

⁶ See especially the first two chapters of his (1994), and compare them to “Appendix B: Meaning and History” of the same book, where he offers his own analysis of Davidson's Swampman: “I think the unbiased intuition is that Swampman thinks all sorts of things that Davidson does: that it's Wednesday, and that radical interpretation is possible, and that water is wet, for example. I think this is because, although he lacks Davidson's causal history, Swampman shares Davidson's dispositions, and it's the counterfactuals that count for content, just as informational theories claim” (1994:117–8). Well, I think, just as his *own* version of informational theory claims. See also Fodor's discussion of what he calls the case of super-Swampman: at the instant you're destroyed and your swampman is created, the same thing happens to your twin. Again, Fodor offers the way his purely informational theory handles them as an argument for it since, he thinks, the intuition that the swampmen in all these cases are genuinely intentional systems at the instant they are created is very strong. This seems like internalism in a very robust sense.

especially given that informational semantics in general has always been taken in the literature as a clear-cut species of externalist theories. But I don't think there is any mystery here. Surely the broad content (denotation/truth-conditions) that the theory assigns to internal mental/brain states are mostly outside the head. But the theory tries to state, in naturalistic terms, what it is about the internal states/mechanisms that makes them *about*, i.e. makes them represent, things that are outside. What the theory does, then, is to account for the naturalistic bases of capacities to represent, be about, those things. (In fact, Fodor's theory, officially at least, does less than that by giving *only* sufficient conditions for intentionality—but put that aside here.) For Fodor, as we've seen, the bases of such capacities are all dispositional, and as such, reside within the head (waiving, again, what he says about information theoretic treatment of proper names and deferential concepts).

The term 'content' ('broad'/'narrow'), very much like 'representation,' is ambiguous: it may denote the thing outside (state, property, whatever) that a brain state is said to represent, be about, or it may denote the *state's having* that semantic property or capacity, which may reside within the head. I believe that Fodor's claim (1990a, 1990b) that meaning (content) is robust is to be understood in this latter sense, which is, I think, intimately connected to his covertly internalist version of informational semantics. According to Fodor, the content of a Mentalese expression has an important degree of independence from the causes of its tokenings, which is to say that an expression is said to have the intentional content or meaning it does even if all the causes of its (past, present, and future) actual tokenings are wild (not occasioned by the "proper" denotation of the expression). Indeed, Fodor's purely counterfactual theory seems to be perfectly apt for naturalizing the mystery of what Brentano has called "intentional inexistence" (1874). It is a naturalistic theory that tries to say, in terms of purely counterfactual covariations, what it is about the internal brain mechanisms that makes certain internal states *be about* things that may or may not actually exist (outside).

So the *de facto* contribution of actual context to the broad content of a symbol drops out in a theory that assigns broad content on the basis of counterfactuals, and as such, PIS *prima facie* assigns disjunctive content to my and my twin's #water# tokens, namely the content (XYZ or H_2O). Fodor wants to resist this conclusion. In a couple of places, he has attempted to show why twins don't have disjunctive broad content given his PIS. I will argue that none of his suggestions works.

3. Fodor's Treatment of Twin Cases

Let's start with what he says in his *The Elm and the Expert* (1994). His official position there *vis-à-vis* the original Twin-Earth example seems to be that twins are impossible since XYZ is a nomological impossibility. If this is granted, Fodor may plausibly assign the broad content H_2O to Earthlings' #water# tokens rather than the disjunctive content (H_2O or XYZ). This move, however, is not effective against Earth-bound twins, and Fodor is aware that it isn't:

However, this brusque treatment doesn't generalize; it depends on the nomological impossibility of XYZ, and I suppose it would be foolhardy to claim that Twin cases are nomologically impossible as such. In fact, I don't claim that they are impossible, or even that they don't happen (cf. the familiar story about jade and jadeite). A broad content psychology would fail to express the generalizations the corresponding narrow content psychology could capture. But I claim that though such cases occur, *it is reasonable to treat them as accidents and to regard the missed generalizations as spurious.* (1994:30-1)⁷

Given this admission, Fodor says that the mental states of such twins have non-disjunctive content, and it's precisely for this reason they can't be subsumed under the same psychological generalizations. But what is Fodor's reason for maintaining that such twins don't have disjunctive content?

⁷ Emphasis in the original. In all the quotations in what follows the emphases will be original unless otherwise stated.

Fodor has some other story (in fact more than one story —see below) to tell as to why in such cases twins have distinct broad contents despite that they share all their dispositions to token their concepts in much the same sorts of circumstances. Given the above quotation, I will restrict my argument to such cases and forget about nomologically impossible twin cases for what follows — as Fodor himself does.

Let's, then, construct an Earth-bound twin case. I'll continue to use XYZ for convenience and assume it to be nomologically possible. The historical situation is pre-modern chemistry, or ancient tribal times, or whatever: people don't have the necessary theory and tools to distinguish between H_2O and XYZ. But they are in fact distinguishable by today's scientific standards perhaps only in high-tech chemistry labs. There are two communities living in different parts of the world. The relations between these two communities are non-existent because different parts of the world are such that people living in one part cannot (i.e. "really" cannot, not just nomologically possibly cannot) travel to the other part of the world. Community A has H_2O around but no XYZ. Community B has XYZ around but no H_2O . According to Fodor, Smith from Community A and twin-Smith from B have, in their Mentalese, tokens of #water# that differ in broad content. According to Fodor, Smith's #water# means H_2O and twin-Smith's XYZ, despite the fact that they seem to share all their internal dispositions to token their #water# symbol.

Question: What makes it the case that the respective tokens of #water# in Smith's and twin-Smith's heads have different broad content?

As far as I could discern, Fodor has tried at least two different answers to this question in different places in his writings. I want to look at them both.⁸

FODOR'S ANSWER #1. Let's start with the answer he gives in a footnote in his (1990a):

⁸ In his (1994:28-33), Fodor appears to give another such answer, but it boils down to the same one I will consider below under the heading "FODOR'S ANSWER #1".

Apparently, then, the content of your term may differ from the content of mine if there's something that prevents tokens of your term from being caused by instantiations of a property whose instantiations could (i.e., really could, not just nomologically possibly could) cause tokenings of mine. This might be true even of two creatures who live in the same world if, as it happens, they live in different parts of the wood. If the nearest XYZ to me is so far away that I can't possibly get there in a lifetime, then, I suppose, "water" means something nondisjunctive in my mouth. Whereas, if the nearest XYZ to you is so close that it's just an accident that you haven't come across any, then I suppose, "water" does mean something disjunctive in yours. (1990a:133:fn.16)

The first thing to notice is that if Fodor wants to stick to this answer, then he has to change his treatment of the broad semantics of concepts like UNICORN, CENTAUR, etc. For our world is such that something really prevents tokenings of #unicorn# or #centaur# from being caused by instantiations of *unicorn* or *centaur*. Even though such properties are nomologically possible (let's grant), there can't be (really can't, not just nomologically possibly can't) any instantiations of them, whatever exactly that means. This passage is from a period when Fodor was experimenting with adding a historical/actual causal interaction clause to his informational account of broad content; so he was in fact considering a different account for UNICORN and the like, an account which takes such concepts to be syntactically non-primitive terms of Mentalese. Perhaps he can make the same move here again.

Indeed he can. But I take it that he wouldn't. For there is a sense in which making this move goes very much against the intuitions that motivate a *purely* informational semantic theory: if having a certain broad content is a matter of having certain dispositions to token a symbol, then what really counts are the counterfactuals. But if Fodor really wants to stick to the above answer then he has to put certain constraints on counterfactuals: he has to categorize them so that only a certain class are allowed to enter into the determination of broad content. We have already seen one such restriction on counterfactuals: counterfactuals that appeal to nomologically impossible situations (relative to our world) are not allowed to

determine broad content; they are excluded from the domain of his purely informational theory. But now with the above answer, he seems to be considering putting some *more restrictions* on them. Well, then, what are they? Can we do any better than just saying “really could, not just nomologically possibly could”?

There are serious difficulties in coming up with a principled criterion, because it seems that anything weaker than nomological impossibility is bound to be relative to the historical, cultural/social, and technological/scientific circumstances that happen to be in place at a given time. Suppose that Community A has stringent taboos against traversing a passage in the mountains that happens to be the only passage connecting the two communities. Even though the route is passable (however difficult and dangerous it might be), no member of Community A dares or will dare to try. So they end up connected to H₂O only. How are we to classify this? Or suppose that members of the two communities can't travel through the wood at a given time because, say, their current scientific/technical means won't allow it, but, say, twenty generations later, they could. What are we to say? Or they can't travel for the same reason, but an alien spaceship comes all of a sudden and suddenly teletransports the members of one community to the other part. Then what? You can generate all sorts of cases that would pump all sorts of bizarre intuitions, and raise all sorts of questions.

My point is not that such questions can have no answer. I suppose you can come up with bizarre answers to bizarre questions in situations like these and then bite the relevant bullets. My point is that if you arrive at a criterion for restricting the counterfactuals that would accommodate all such scenarios, it will likely be arbitrary and *ad hoc*, and will answer and satisfy no important intuitions.

I don't suppose that Fodor will want to have anything to do with this mess. Instead, I hope, he will be tempted to say “yuck!” What he ought to say, sticking to his pure informational semantics, is that until they discover the true essence of their respective stuffs, they have disjunctive content. Because learning what something really is changes your

dispositions.⁹ You cease to be the same function from contexts to broad contents, to put it in the jargon of narrow content — see below.

FODOR'S ANSWER #2. I suppose Fodor's official answer to why twins don't have disjunctive broad content is to be found in his (1990a). Here is what it comes down to:

... "water" means water (and not XYZ) because, although people would use "water" of XYZ if there were any (XYZ is supposed to be indistinguishable from H₂O) nevertheless, they have a settled policy of using "water" as a kind-term (of using it only for substances actually of the same kind as water), and their adherence to this policy makes their use of "water" for XYZ asymmetrically dependent on their use of "water" for H₂O: there's a break in the XYZ/"water" connection *without* a break in the H₂O/"water" connection in nearby world where H₂O is distinguishable from XYZ. (1990a:116)

Notice that, as stated, this is question-begging as far as naturalism is concerned. You can't use intentional terminology (like having a settled policy of using "water" as a kind term) in an attempt to state what makes "water" mean what it does. Here Fodor seems to be characterizing what makes our use of *linguistic* symbol "water" mean what it does. Our concern, however, is with the semantics of Mentalese. But when you transfer the story to apply there, things are different. On pain of circularity, you can't appeal to any policies, intentions, etc. in the tokening of #water#. Fodor is aware of this, but what he says in a footnote is cryptic:

I take it that, but for the talk about intentions and policies, the same sort of line applies to kind-concepts. What makes something a kind-concept, according to this view, is what it tracks in worlds where instances of the kind to which it applies are distinguishable from instances of the kinds to which it doesn't. (1990a:134:fn.25).

⁹ Cf. Fodor (1987:94-5) and (1991:302-4, Reply to Stalnaker).

Now the first thing to say about this is that the same sort of line does *not* apply to kind-concepts! This is obvious from what he says in the next sentence. What he says there about what makes a Mentalese symbol a kind-concept is different from having intentions, policies, and so on. It may be that you're quantifying over intentional mechanisms like having policies, with the result that the symbol should track instances of what it applies to. That is fine, but this formulation crucially differs from talk of intentions and policies. The reason I am insisting on this is that Fodor's answer to our main question must be stated in terms of this "tracking" formulation about what makes a symbol a kind-concept. If so, however, his answer doesn't work.

Twins are molecularly identical. So all their dispositions are identical. So, *a fortiori*, all their dispositions to apply their #water# symbol are identical. And, as a matter of fact, they do or would apply it to either stuff.

Question: What makes twin-Jones' #water# symbol to mean XYZ but not H₂O despite the fact that he would apply it to H₂O?

Fodor's answer must be that twin-Jones' present disposition to apply #water# to H₂O asymmetrically depends on his present disposition to apply it to XYZ. (The same question and answer, *mutatis mutandis*, go for Jones.) But is that true?

Notice that Fodor's answer to the question of what underwrites this asymmetric dependence when translated to "tracking" terminology comes to this: relative to twin-Jones' present time and world, there are worlds in which he applies #water# to XYZ but not to H₂O where the two substances are distinguishable, i.e. there are worlds in which his #water# symbol tracks XYZ but not H₂O, and there are also worlds in which he applies #water# to both substances where they are indistinguishable.¹⁰

But these worlds are irrelevant to answering the above question: viz., what makes twin-Jones' *present disposition* to apply #water# to H₂O asymmetrically dependent on his

¹⁰ Cf. Fodor (1990a:115–6) where he elaborates on his answer in more detail — comparing the present situation with the way he treats the semantics of UNICORN and with an objection by Baker about cats and robot-cats.

present disposition to apply it to XYZ? This is the question that must be answered if Fodor's asymmetric dependence claim is to justify his attempt to assign non-disjunctive content to twins. In order to answer *this* question, however, you have to consider worlds in which twin-Jones' present dispositions are kept the same. Remember that the asymmetric dependencies must be synchronic. In other words, it is irrelevant how you have gotten or developed your present dispositions in the first place, i.e., on the basis of what previous dispositions you had.¹¹ Similarly, how your present dispositions *will* change/evolve is irrelevant to determining the objective dependencies among your *present* dispositions. But in the worlds Fodor considers, the dispositions are no longer the same. All the worlds in which twin-Jones' #water# symbol tracks XYZ but not H₂O are worlds in which H₂O and XYZ are distinguishable by twin-Jones. But in these worlds twin-Jones presumably has a whole new set of dispositions resulting from learning a chemical theory that says what XYZ and H₂O are and how to tell them apart. As Fodor says, learning what a thing really is changes your dispositions to apply your concepts, hence the extensions of your concepts. These worlds, therefore, can't be relevant to evaluating the claimed asymmetric dependence claim. In the worlds twin-Jones can tell XYZ and H₂O apart he has a significantly different set of dispositions than the one he has in worlds he can't. And these differences will make twin-Jones no longer a twin of Jones, viz., Jones and twin-Jones in their parallel histories will cease to be molecular duplicates in ways that would affect our evaluation of asymmetric dependence claims for both.

So, what are the counterfactuals relevant to assessing the claim that whereas twin-Jones' current disposition to apply #water# to H₂O asymmetrically depends on his disposition to apply it to XYZ, Jones' current disposition to apply #water# to XYZ asymmetrically depends on his disposition to apply it to H₂O? In all the relevant counterfactual worlds, Jones and twin-Jones must not cease to be molecular duplicates, i.e., they must have the same set of dispositions they now have. That is because we are

¹¹ Cf. Fodor (1987:109) and (1991, Reply to Baker).

considering the *synchronic* dependencies of their *present* dispositions, i.e. we are trying to figure out which dispositions are *now* dependent on which. This is the question that must be answered. When the question is put in this way, however, the answer is obvious: Jones' present dispositions to apply #water# to both substances *symmetrically* depend on each other, and similarly for twin-Jones. But this means that the broad content of their #water# tokens is disjunctive, i.e. they express *H₂O or XYZ!*

I conclude that Fodor's PIS assigns disjunctive content to twins. More generally, it assigns broad content to symbols *independently of any particular contexts*, or equivalently, *without regard to what particular context the subject happens to be actually in*. But this was in effect the basic idea underlying Fodor's notion (1987:Ch.2) of narrow content he developed essentially as a response to twin-earth cases.

The narrow content of a primitive Mentalese expression token is, according to Fodor, a set of ordered pairs. The first element in the pair is a context and the second the broad content (denotation/truth-conditions) the expression token would have in that context. The narrow content is therefore a partial function or mapping from contexts to broad contents. What twins share is the narrow content of their mental states since they instantiate the same function from contexts to broad contents. Here is how Fodor makes much the same point:

Take my (syntactically individuated) 'dog' concept *together with its associated covariation-causing mechanisms* to Twin-Earth, and what you get is 'dog'/twin-dog covariation, instead of the 'dog'/dog covariation you get around here. In effect, as you carry the 'dog'-and-mechanism pair from world to world, it picks out a set of properties; one for each world in which the narrow content of 'dog' is defined. This set of worlds-and-properties (including *Earth/dogness, Twin-Earth/twin-dogness, etc.*) is the narrow content of my mental symbol 'dog'. (1991:269, Reply to Block)

Despite the general agreement that narrow content is supposed to be in the head, according to this passage, Fodor's narrow content ain't in the head! For the set of ordered pairs in question are obviously not in the head. We need to make a distinction here between the narrow

psychological state, viz. a symbol-and-mechanism pair itself, and the narrow content, viz. the set of <context, broad content> pairs, that gets assigned to the narrow state. Although the former is in the head, the latter is surely not. This is the sense in which Fodor's notion of narrow content may be said to be extensionalist/externalist, since the individuation of narrow psychological states on the basis of their narrow content is extensionalist/externalist.

When Fodor talks about narrow content's being in the head,¹² what he means to be talking about, probably, is that the narrow state, whose individuation conditions are given extensionally/externally, is in the head.¹³ The narrow internal state that gets individuated by such a notion of narrow content is in fact an equivalence class of particular symbol-and-mechanism pairs. To put it intuitively, the narrow content of a mental state picks up a certain set of internal dispositions supervenient on the brain that would cause the mental state covary with a range of external stuff in different contexts. The point about Fodor's mapping notion of narrow content is that the individuation of these internal dispositions is done externally. Any mechanism/syntactic-object pair that would effect the same mapping from contexts to broad content would have the same narrow content.

But, when put in this way, it's clear that Fodor's PIS turns out to be (more or less)¹⁴ extensionally equivalent to his previous notion of narrow content, since the broad content of a syntactic object given a mechanism is determined by what it would nomically covary with in different contexts. The broad content, then, is the disjunctive sum-total of properties in

¹² E.g. Chapter 2 of *Psychosemantics* (1987).

¹³ If this is right, then, contrary to what Fodor says in Chapter 2 of (1987), narrow content turns out to be expressible in principle after all. As I said above, the term 'content' ('broad'/'narrow'), very much like 'representation,' is ambiguous: it may denote the thing outside (state, property, whatever) that a brain state is said to represent, be about, or it may denote the *state's having* that semantic property or capacity, which may reside within the head. The set of ordered pairs that is said to be the narrow content (first sense) of a token doesn't reside in the head; it is only the narrow content in the second sense that supervenes on what is inside the head: this is the token's *having* the narrow semantic capacity, i.e. whatever is with the internal grounds or bases (covarying symbol-and-mechanism pairs) that makes the token be assigned the set of ordered pairs (=narrow content) in question. Similarly for 'broad content'. Although I point out the ambiguity here, in what follows I'll be largely ignoring the distinction I've just drawn, leaving to the context the job of disambiguating 'content' or 'meaning.'

¹⁴ More or less, because we've accepted that the domain of PIS must be restricted to worlds that are nomologically accessible from our world.

different contexts with which the mechanism-symbol pair nomically covaries, just as his PIS requires.¹⁵

Ironically, Fodor claims to have changed his mind on narrow content in his *The Elm and the Expert* (1994): he says that intentional explanation requires only broad content and that therefore we can dispense with narrow content. But, alas, as we have seen, the notion of broad content his PIS delivers is extensionally equivalent to his previous notion of narrow content.

Fodor's asymmetric dependence doesn't work. It doesn't block the consequence that twins have non-disjunctive content. Fodor (in personal communication) has indicated that he intends to elaborate his response along the answer I named 'FODOR'S ANSWER #1' above. I am not sure I got his reply right. Here is a sketch of his reply. Among the counterfactuals that support the law "water #water#," only some are relevant to the truth of the law. Others are to be discarded. Fodor refuses to say which are which. He seems to think that he is not obliged to give a criterion. He thinks that the notion of a *local law* is all he needs and that trying to cash out what a law is in terms of counterfactuals is a bad idea anyway. Moreover, he thinks that informational semantics has always been in need of a notion of local law that is committed to making this distinction among counterfactuals.

¹⁵ More strictly, since the narrow content of a Mentalese expression, #F#, is officially specified as a set of ordered pairs, $\langle w, d \rangle$, whose first element (w) is a context and the second a broad content (denotation) #F# would have in that context, Fodor needs to make it sure that his pure informational semantics assigns the right, i.e. the intended, denotation to #F# in each particular context, w . But his informational semantics, being purely counterfactually stated assigns to #F# a disjunctive content, i.e. a disjunction of denotations comprised, intuitively, by what in fact is intended to be the second elements of all the ordered pairs in the set (i.e., the set to be identified as the narrow content of #F#). Furthermore, it does assign the very same disjunction in every context, i.e. for each and every first element, w , of the ordered pairs in the same set. Thus #F# ends up having the same broad content in every contexts, namely a disjunctive one. So, for instance, since the intended broad content assignment for my twin's #water# token is XYZ, for mine H_2O , etc., the narrow content of our #water# tokens is actually meant by Fodor to be $\{ \langle \text{Earth}, H_2O \rangle, \langle \text{Twin-Earth}, XYZ \rangle, \dots \}$. But his informational semantics assigns the disjunctive broad content $(H_2O \vee XYZ \vee \dots)$ to our respective #water# tokens. If so, the narrow content of our #water# tokens is $\{ \langle \text{Earth}, (H_2O \vee XYZ \vee \dots) \rangle, \langle \text{Twin-Earth}, (H_2O \vee XYZ \vee \dots) \rangle, \dots \}$. So the narrow content of #water# as a partial function from contexts to broad contents is a constant function: it gives the same value (broad content) for every argument (context). But this seems to be exactly what Fodor's informational semantics does too: it assigns broad content irrespective of any *actual* contexts.

I am not sure how to take this suggestion in the absence of a more detailed elaboration. But I think Fodor's reply seems to reflect something important about the notion of information that his own account ignored. So let me reflect on what information was supposed to be after all. I think there are interesting issues to pursue here in connection with Fodor's reply. In particular, let me discuss the notion of informational content of twins' relevant mental states. (This is, at any rate, necessary before taking up the notion of intentional or semantic content which is supposed to be built upon the notion of information.) So what is the informational content of my and my twin's "water"-thoughts? To answer this question, we need to know what information is in the relevant sense. It seems to me that we don't have an adequate notion of informational content unproblematic and precise enough to guide us through some of the different and controversial cases like Earth-bound twins. As advertised, in what follows, I'll paint a picture in broad brushes about how a pure informational theory would perhaps best handle twin cases accommodating both internalist and externalist intuitions.¹⁶

4. Information and the Broad/Narrow Dichotomy: Sketch of a Picture

Dretske's 1981 account gives the following characterization:

A signal r carries the information that s is F = The conditional probability of s 's being F , given r (and k), is 1 (but, given k alone, less than 1). (1981:65)

As many noted, there are several problems with this characterization that relies on inverse conditional probabilities whose theoretical cogency is moot. Even Dretske seems to have dropped this way of characterizing information. But there are nevertheless several aspects of the characterization that need to be highlighted. It makes it clear, for instance, that there is no such thing as misinformation. Necessarily, if r 's being G carries the information that s is F then s is F . What seems to be essential is that there is a lawlike relation between the two events

¹⁶ For a more detailed and worked out account of what is to follow, see my (in prep.-b).

given certain channel conditions that obtain between them. In other words, the two events must be connected to each other through a series of nomic dependencies. Let's try to work out some of the details on a familiar example before taking up twins.

Consider Dretske's own example in his "Misrepresentation" (1986). Certain marine bacteria have little internal magnets (called magnetosomes) that align them parallel to the earth's magnetic field. Since magnetic lines in the northern hemisphere incline downwards, the bacteria propel themselves away from oxygen rich surface water, considered toxic for the bacteria, towards geomagnetic north, i.e. towards oxygen-free sediment at the bottom. The survival value of magnetotaxis, as this sensory mechanism is called, must be obvious. The question may be put this way, what is the information content of the orientation of the magnetosomes when they orient the bacteria in the direction of geomagnetic north? Is it that *that* is the direction of geomagnetic north? Or, *that* is the direction of oxygen-free water? Dretske's own answer seems to be the former, since he thinks that the lawful covariation is between the particular orientation of the magnetism and the direction of geomagnetic north. Even though the direction of the geomagnetic north in that particular environment just happens to be the direction of the oxygen-free water, the information content of the magnet's orientation is: *that* is the direction of magnetic north, and not: *that* is the direction of oxygen-free water. But how do we tell what does nomically covary with what? Well, you appeal to counterfactuals: if you put the bacteria in the southern hemisphere, their magnetosomes would orient them in the direction of geomagnetic north, not in the direction of oxygen-free water. In fact, in southern hemisphere, the direction of geomagnetic north is the direction of oxygen-rich, toxic, surface water. So if you plant the northern bacteria in the south, they would destroy themselves. The magnetosomes, in other words, track geomagnetic north in the relevant counterfactual situations where they work properly. They don't track the direction of oxygen-free water. The dependency between the directions of geomagnetic north and the oxygen-free water is accidental, not nomic. It occurs only in some environments. So we may conclude along with Dretske that the information the states of magnetosomes carry is only about the

direction of the magnetic north.¹⁷ In general we may take the inner mechanisms of bacteria as the relevant unit to be kept intact and may change pretty much everything else outside, i.e. put them in any environments we like, in determining the informational content of their magnetotaxic states, hence we may decide to be very liberal about our counterfactuals. In other words, the channel conditions that are supposed to be kept intact in fixing the relevant information content are to be restricted to the inner mechanisms of the organisms.

As we've seen, Fodor's version of PIS, at least as it occurs in his (1987) and (1990a), seems to take the inner perceptual/cognitive mechanisms of an intact organism as the main channel conditions to be kept fixed in determining the informational content of the states of the organisms. Almost everything else outside the skin of the organisms can be counterfactually varied, so much so that, the non-existence of, for instance, unicorns, centaurs, etc. in the actual environment of organisms becomes irrelevant in fixing what laws exist in the determination of informational content.¹⁸ Here the information-making laws are pretty much non-local: counterfactuals are allowed to vary over a wild range of possibilities even if we restrict them to only nomological possibilities. Call this approach the *liberal or internalist interpretation of informational content*, since it abstracts from the local contingencies of actual environments and assigns disjunctive content to a given state-type of the organism depending on what it would reliably covary with given only the internal dispositions of the organism. On this interpretation, the informational content of my water-thought is disjunctive: roughly, it consists of the disjunction of all the different stuffs I cannot discriminate between.

¹⁷ This is Dretske's position in (1986). In his book (1987), he seems to have moved closer to Millikan's (1984, 1989) views in handling the problem of misrepresentation. But cf. his 1994. In "The Epistemology of Belief" (1983), he takes again what seems to be a different line. This is one of those rare occasions where he explicitly discusses the Twin-Earth cases. And the solution he comes up with is somewhat similar to and in certain ways better articulated than Fodor's own suggestions about the locality of laws adverted to in fixing the informational content.

¹⁸ There are exceptions to this liberal attitude in Fodor's writing, especially when he wants to include the existence of experts and actual baptizing events as parts of the mechanisms (channel conditions) sustaining the information-making laws. I want to leave this issue aside however for the purposes of this paper.

Notice that this approach need not include among the disjuncts stuff like vodka-in-a-clear-glass, that only accidentally cause the tokening of my “water” thoughts, and only certain occasions. There are a variety of theoretical devices one might want one to use to exclude these latter cases: Fodor appeals to asymmetric dependence among the internal dispositions. Another solution perhaps would be to appeal to optimality conditions. Drestke’s own solution is to appeal to a learning period during which no mistakes are allowed.

I think that this way of looking at the information is useful, although, I admit, a little idiosyncratic. What makes it useful is that the channel conditions that are to be fixed before one starts talking about information are all part of the inner mechanisms or dispositions of cognitive organisms, hence it is responsive to the internalist/individualist intuitions. Instead of talking about narrow content which is supposed to supervene totally on the inner make-up of an organism, we can usefully talk about disjunctive informational (broad) content which supervenes on what is inside the skin. This seems like internalism in a very robust sense. If so, what we have here is an internalist theory of broad content exclusively on the basis of what information is carried by the internal states (assuming that we have an independent story to tell how error/misrepresentation is possible). This may sound paradoxical, but actually it is not. As I said above, the broad content (denotation/truth-conditions) that the liberal informational approach assigns to internal mental/brain states are outside the head. But the approach tries to state, in naturalistic terms, what it is about the internal states/mechanisms that makes them carry information *about* things that are outside the organism; if we can handle the misrepresentation problem, we have the beginning of a naturalistic story about what makes these states represent things that are outside. What this approach does, then, is to provide the beginnings of an account for the naturalistic bases of capacities to represent, be about, those things.

But there are other, more specific, ways of interpreting information: we may be more inclusive in specifying the channel conditions that decide what information is carried. Let’s go back to magnetosomic bacteria for a moment. Why can we not say, for instance, that the *local*

laws in the northern hemisphere are such that there is after all nomic covariation between the direction of the magnetic north and the oxygen-free water? We just have to restrict our counterfactuals carefully: in particular, we have to restrict them to Northern Hemisphere. Consider: if this bacterium were to be taken from the Aegean Sea and put in the Norwegian sea, the orientation of its magnet would be in the direction of oxygen-free water. Even with this geographic restriction, then, there are indefinitely many such counterfactuals to support. We certainly have the counterfactuals' power here. It is just that they have to be confined to certain times and places. Local laws state what local dependencies there are. And these dependencies are nomic, i.e. lawlike given that the counterfactuals supporting them are to be locally restricted. So the information that the states of magnetosomes carry in the Northern Hemisphere is after all about the direction of oxygen-free water, not about the magnetic north. Call this approach the *restricted or externalist interpretation of information*, since the counterfactuals are to be restricted according to what local contingencies there are in the actual environments. Here the channel conditions go out of organisms and reach out the world, so to speak. There is, of course, still room for variations about which local contingencies are to be respected and to what degree.

Similarly, given that there is in fact no XYZ around in this world, our #water# tokens carry information about H₂O even though they *would* covary with XYZ if it *did* exist here. Local laws ought to be restricted to counterfactuals that advert to aspects of our actual world, or to its very close cousins, where there is only H₂O around.

This restricted or more externalist approach to interpretation, also, is very attractive since it accommodates some of the externalist intuitions. In fact it is closer to the account Dretske originally developed in his *Knowledge and the Flow and Information* (1981). It is, of course, also problematic in the same ways that Dretske's original account was problematic in its treatment or characterization of channel conditions and the assumed background knowledge expressed by the constant *k* as I quoted above. Information seems to be a matter of what nomic relations obtain between the source and the receiver given a certain set of stable

channel conditions. What makes certain conditions channel conditions is that they don't generate information: they are fixed relative to a given framework. The information is supposed to be generated by what possibilities exist at the source. But when put this way, it is not clear whether the actual non-existence of XYZ is supposed to be treated as part of the channel conditions or as a relevant possibility at the "source". The indeterminacy is quite general. Take again the case of the magnetotactic bacteria. If being in the Northern Hemisphere is taken as part of the channel conditions, then the relevant possibilities will include the direction of oxygen-rich or oxygen-free water: hence the relevant information carried is about oxygen-rich or oxygen-free water. If the channel conditions are restricted only to the stable/proper workings of the bacteria's inner mechanisms, then the possibilities that the states of the magnetosomes reliably covary with will only be the features of the magnetic field.

Once you go beyond the skin of a cognitive organism and begin citing external factors in the specification of channel conditions, what information is carried will be relativized according to your choice. And your choice will be guided by what particular explanatory/predictive/practical interests and purposes you may have. This was exactly my point in criticizing what I called above FODOR'S ANSWER #1. But now I would like to suggest that this relativization may not necessarily be a bad thing. I think our externalist intuitions demand that we should so relativize the notion of information for certain purposes and interests.

I think it may be a mistake to think that we have a unique and determinate notion of semantic content that answers only one set of determinate purposes and interests. The kind of (disjunctive) content that the liberal/internalist interpretation of information delivers is only one notion of content that is perhaps more relevant in the contexts of psychological explanation and prediction. It would accommodate twins, replicas, Swampmen and the like. And it is apt for replacing the traditional notion of narrow content, and solve Brentano's problem and answer the challenge it poses to the naturalist: what is it about human minds

that are capable of entertaining thoughts about things that don't exist? In other words, how is "intentional inexistence" possible?

On the other hand, the kind of content(s) that the restricted/externalist interpretation of information delivers is more relevant in contexts where the issue is what guides us to the world in which we exist, find our way through and try to survive, or even where the issue is what local contingencies have determined the selectional forces in our biological or cognitive development. That there may be no one principled answer to all the relevant questions I raised in my scenario about the earth-bound twin communities above may not perhaps be so troubling when viewed in this light. The notion of information that the restricted/externalist interpretation delivers is certainly well defined even though there may be no unique interest- or purpose-free way of deciding what conditions will count as channel conditions in a given case. This doesn't make information any less subjective or less naturalistically acceptable.

I would like to leave my discussion here at this intuitive and very sketchy level. In a way, I would like to have it both ways if I can: I would like to have both narrow and broad content, only that the former isn't really narrow in the usually intended sense in the literature. An information based semantics may be capable of delivering all the notions of content that we need in accommodating both internalist and externalist intuitions.

REFERENCES:

- Aydede, Murat (in prep.-a). "Dretske's Informational Semantics."
- Aydede, Murat (in prep.-b). "From Information to Intentionality."
- Devitt, Michael (1996). *Coming to our Senses: A Naturalistic Program for Semantic Localism*, Cambridge University Press.
- Dretske, Fred (1981). *Knowledge and the Flow of Information*, The MIT Press.
- Dretske, Fred (1983). "The Epistemology of Belief," *Synthese* 55, 3–19.
- Dretske, Fred (1986). "Misrepresentation" in *Belief: Form, Content and Function*, (Radu J. Bogdan (ed.), Clarendon Press, Oxford, 1986.
- Dretske, Fred (1987), *Explaining Behavior*, The MIT Press.
- Dretske, Fred (1994). *Naturalizing the Mind*, The MIT Press.
- Fodor, Jerry A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, The MIT Press.
- Fodor, Jerry A. (1990a). "A Theory of Content, II" in *A Theory of Content and Other Essays*, The MIT Press.
- Fodor, Jerry A. (1990b). "Information and Representation" in *Information, Language, and Cognition*, Philip P. Hanson (ed.), Vancouver: University of British Columbia Press, 1990.
- Fodor, Jerry A. (1991). "Replies" (Ch.15) in *Meaning in Mind: Fodor and his Critics*, B. Loewer and G. Rey (eds.), Basil Blackwell, 1991.
- Fodor, Jerry A. (1994). *The Elm and the Expert: Mentalese and Its Semantics*, The MIT Press.
- Loewer, Barry and Georges Rey (eds.) (1991). *Meaning in Mind: Fodor and his Critics*, Oxford: Basil Blackwell.
- Millikan, Ruth G. (1984). *Language, Thought, and Other Biobiological Categories*, The MIT Press.
- Millikan, Ruth G. (1989). "Biosemantics," *Journal of Philosophy* 86, no.6, 281–297.