Mind & Language, ISSN: 0268-1064
Vol. 12. Nos. 3/4 September/December 1997, pp 422–458.

Has Fodor Really Changed His Mind on Narrow Content?

MURAT AYDEDE

Abstract: In *The Elm and the Expert* (1994), Fodor rejects the notion of narrow content as superfluous. He envisions a scientific intentional psychology that adverts only to broad content properties in its explanations. I show that there has been no change in Fodor's treatment of Frege cases and cases involving the so-called deferential concepts. And for good reason: his notion of narrow content (1985–91) couldn't explain them. The only apparent change concerns his treatment of Twin Earth cases. However, I argue that the notion of broad content that his purely informational semantics delivers is, in some interesting sense, equivalent to the mapping notion of narrow content he officially gave up. I also critically reconstruct the evolution of Fodor's thinking between 1980 and 1994.

1. Introduction

Here is what happened. Fodor wanted to give a naturalistic account of intentionality. Since he assumed a computational picture of the cognitive mind, his task was to give an account, in non-intentional/non-semantic terms, of what it is for non-logical primitive expressions of Mentalese to represent what they do. He thought that functionalist considerations can't play any role in any part of such a naturalistic account, since for Fodor functionalism in semantics either yields a destructive sort of holism or otherwise is committed to an already defunct analytic/synthetic distinction. This meant, for Fodor, that meaning is purely denotational and truth-conditional. And the job of a naturalistic theory of meaning, therefore, was to account, in non-semantic terms, for what it is for a Mentalese primitive token to have the denotation or truth-conditions that it does. Fodor opted for a purely informational story as the naturalistic story to be told. I will come to it in more detail below, but for the moment, the core idea of the theory is roughly that

I would like to thank Marshall Abrams, Jon Cohen, Ed Cox, Michael Devitt, Jerry Fodor, John Kulvicki, Ronald Loeffler, Ruth Millikan, Jesse Prinz, Philip Robbins, Carol Slater, Robert Wilson and an anonymous referee of this journal for their help, encouragement and/or comments on earlier versions of this paper.
Address for correspondence: Department of Philosophy, The University of Chicago, 1010 East 59th Street, Chicago, IL 60637, USA.
Email: m-aydede@uchicago.edu.

the representational, hence denotational, content of a primitive Mentalese token reduces to what it would nomically covary with.

Along the way, however, Fodor has been conceiving of psychology as an intentional discipline, that is, a discipline whose laws advert to intentional properties of mental states. Psychological explanation was, for Fodor, a matter of subsuming the explananda under intentional laws; in other words, psychological phenomena are covered by laws *qua* intentional. This meant that psychological events are subsumed by laws in virtue of their denotational or truth-conditional properties.

On the other hand, Fodor thought that Twin Earth cases somehow show that intentional psychology can't be adverting to denotational content. At this point three constraints needed to be satisfied:

(1) keep psychology intentional,
(2) make intentional properties that are adverted to in psychological laws non-denotational, and
(3) keep (semantic) functionalism out.

The result was his notion of narrow content, which is explicated, as we will see in more detail in a moment, in *extensional/externalist* terms as a partial function (or mapping) from contexts to denotation/truth-conditions (broad content), even though it was supposed to supervene, in some sense, on what is inside the head. This was more or less Fodor's thinking between approximately 1985 and 1991.[1] Before this period, his notion of narrow content differed in that it was not hostile to functional role.[2]

After that period, it is fairly common to think that Fodor's views on narrow content and its relation to intentional explanation underwent quite radical changes; in particular, that he dropped the notion of narrow content altogether as a superfluous construct after around 1992. His first published remarks on this change came out in *The Elm and the Expert* (1994).

I will argue that, despite the appearances to the contrary, Fodor's views haven't really changed much, certainly not in any way that would warrant calling the change 'radical'. What he did was just to fine-tune and readjust some details, mostly by redrawing/relabelling the theoretical space without substantially changing it. I will then show that, given the intuitions that have

---

[1] Roughly, from 'Fodor's Guide to Mental Representation' (1985) and *Psychosemantics* (1987) to 'A Modal Argument for Narrow Content' (1991b). See also his 1986.

[2] See his 1981 and 'Responses' to peer commentary and his (unpublished). In his 1982, Fodor attempts to give a phenomenological characterization of narrow content using the same apparatus of a partial function from worlds to broad contents. Interestingly, this apparatus makes him relatively immune to holistic consequences of a functionalist characterization. In this regard, it is to be compared to his (unpublished) where he proposes again the phenomenological content as the narrow content, but he attempts to characterize it this time in terms of the Ramsey sentence of phenomenological generalizations, and not surprisingly, he starts worrying about holism and indeterminacy of narrow content.

been driving Fodor's thought all these years, there is a better way of redrawing the space from his own perspective.

The paper has mainly two threads. One is historical, exegetical and reconstructive: I trace the evolution of Fodor's thought over the past 12 years about (psycho)semantics and psychological explanation. Although it is exegetical, its primary aim is critical: I try to give the reader a coherent overall sense and (mostly reconstructed) picture of Fodor's thought and its development, its underlying unity, its merits, as well as some of its inherent difficulties and pitfalls. While this thread is the thicker one, the other thread is more substantive but sketchy: it will ultimately suggest a picture according to which a purely informational semantics à la Fodor can be seen as a species of an internalist theory! The details of this thread will become apparent as I start presenting and criticizing Fodor's previous 'mapping' notion of narrow content *vis-à-vis* the notion of broad content that his own version of informational semantics delivers (especially in the context of Twin Earth cases). I will argue that there is a way of seeing that these two notions collapse into one. I admit that this sounds odd, but I believe this is what makes this thread by far the most interesting. But we will need the other, more exegetical, thread to prepare the way. So I hope that the paper has a broader scope and interest than just being an exegetical piece on Fodor in that if one is interested in a pure informational semantics like Fodor's, one may not perhaps be so off-target in accommodating internalist intuitions without introducing any notion of narrow content. Conversely, if one is interested in a mapping notion of narrow content like Fodor's, one doesn't perhaps necessarily occupy a position that is problematic *vis-à-vis* externalist intuitions.

Since this is a rather long paper, let me give a somewhat detailed outline of what is to come. Since I want to show that there is a sense in which Fodor hasn't really changed his mind on narrow content, I will have to examine his notion of narrow content at some length. This is the job of Section 2, but it will occupy us in other sections as well. In Section 3, I want to show two things: first, that Fodor's notion of narrow content had nothing to do with Frege cases, and second, that Fodor's (independent) treatment of Frege cases was (and still is) problematic. My discussion here is an attempt to critically reconstruct Fodor's thinking between 1985 and 1991 *vis-à-vis* Frege cases. This is a slight digression, but it will be important since I want to argue that Fodor's later rejection of narrow content, contrary to his own advertisement in his (1994), had very little effect on his vision of how to handle Frege cases, and generally, on his vision of a scientific psychology. Section 4 will examine the alleged consequences of Fodor's rejection of narrow content. There I will follow Fodor's own strategy. In his (1994), where he rejects narrow content, he goes over three cases whose psychological explanation is usually thought to motivate a notion of narrow content. These are (1) Frege cases, (2) cases of so-called deferential concepts, whose semantics seems to require experts' knowledge, and (3) Twin Earth cases. All of these are supposed to motivate construction of a notion of narrow content. After briefly discussing the first

two in Sections 4.1–4.2 and showing that Fodor's position didn't change a bit, I will turn in Section 4.3 to Twin Earth cases.

Fodor's notion of narrow content as a mapping from contexts to contents was first developed in the second chapter of *Psychosemantics* (1987), and was designed primarily to handle the twin cases. Fodor's views on twins indeed appear to have undergone radical changes. But I will argue that the appearance is misleading. I will show that his notion of narrow content collapses into the notion of broad content which results from the pure informational semantics he first developed in the fourth chapter of the same book (1987). In particular, I will argue that Fodor's pure informational semantics assigns disjunctive broad content to twins, since, being a purely counterfactually stated theory, it assigns broad content to mental states without regard to the actual contexts organisms happen to be in. I will take up three different attempts by Fodor to block this obvious consequence of his theory and show that none of them works. In a nutshell, then, I will argue that there wasn't much of a difference between Fodor's notion of broad and narrow content to begin with. Hence, when Fodor rejects *his* notion of narrow content, all the explanatory apparatus he previously used is still (almost entirely) in place in his 'new' theoretical position. I will also suggest how the notion of broad content Fodor's purely informational account delivers may be capable of taking over all the important jobs his notion of narrow content was supposed to do. So it will turn out that Fodor's old and new positions are extensionally equivalent. The more substantive part of my discussion will therefore emerge from my discussion in Section 4.3.

2. *Fodor's Notion of Narrow Content*

To begin with, let's see the way Putnam's Twin Earth example was supposed to motivate Fodor's mapping notion of narrow content. The details will be important later on.

The intuitions to which Putnam's thought experiment appeals concern the alleged essential contribution of context to the determination of the intentional content of mental states. Since twins are supposed to be molecularly identical, what is inside their heads is type-identical, but the intuitions allegedly demand that twins' corresponding mental states are about different stuff: mine, say, is about H_2O, but my twin's has XYZ as its representational content. So if intentional properties are all denotational (broad), twins are in different mental states despite their psychological identity. But then, if the laws of psychology advert to broad content, twins are not subsumed by the same laws. This result, Fodor thought (among many others), is intolerable in that it flouts the mind–body supervenience and offends against the quite robust intuition that twins are interestingly similar in their behaviour and not accidentally so: for all intents and purposes relevant to psychological explanation, twins' mental states form a natural kind (they have the same causal powers across contexts), and therefore they should be subsumed by

the same generalizations.[3] Hence psychological laws can't be adverting to broad semantic properties of mental states.

Fodor said, 'let there be another kind of content then',[4] one that supervenes on the agent's physiology, a narrow content whose identity conditions are nevertheless given extensionally and externally, *a fortiori* not (narrow) functionally. The narrow content of a primitive Mentalese expression token is, on this conception, a set of ordered pairs. The first element in each pair is a context and the second the broad content (denotation/truth-conditions) the expression token has 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 because they instantiate the same partial function from contexts to broad contents.

Sometimes Fodor characterizes narrow content as a mapping from thoughts *and* contexts to broad contents.[5] But most of the time he uses the formulation I gave above as a mapping from contexts only. I think that the former formulation is untenable for at least the following reason. Narrow content is supposed to be a construct that essentially figures in psychological generalizations. In other words, narrow contents are, essentially, supposed to be interpersonally shared and ascribable. But if we specify narrow content as a mapping from thoughts and contexts, we risk the danger of losing this aspect of narrow content unless we also assume some sort of a type-physicalism, since the characterization of 'thoughts' here can only be 'syntactic', i.e. a non-semantic characterization of the Mentalese vehicle that is said to have the narrow content in question.[6] In the case of twins, this is not a problem because twins are physically type-identical by stipulation. Not only are their #water# tokens type-identical (physically as well as functionally), so are all the psychological/physiological mechanisms that sustain the covariation between their #water# tokens and the different stuff in their respective environments (H_2O in my context; XYZ in my twin's). Physical identity of organisms/systems is a sufficient condition for sharing narrow content. But if the notion of narrow content is going to be of any use in psychology, it would be wise to have an identity criterion for narrow contents that doesn't assume underlying physical identity of the organisms to

---

[3] This way of putting the matter begs the question against anti-individualists like Burge 1986. But I will ignore the complications since my concern here is with how Fodor himself conceived of the dialectics of the problem space.

[4] As Fodor himself acknowledges, his account of narrow content exploits the notion of character in Kaplan's 1989 account of the semantics of demonstratives. For a similar notion of narrow content, see White 1982.

[5] E.g. Fodor 1987, pp. 47–8.

[6] Despite Fodor's recent official functionalist position (e.g. 1994, pp. 106–9), I am assuming here that interpersonally applicable narrow functional individuation of the primitive non-logical symbol tokens of Mentalese is out of the question, for reasons exactly parallel to the ones Fodor himself gives when he rejects functional role semantics. I will say more on this below, but for a detailed elaboration of this claim, see my 1996.

whom such contents are attributed, since twins only occur in philosophers' imaginations.[7] This is more or less acknowledged by Fodor:

> This set of worlds-and-properties (including *Earth/dogness, Twin-Earth/twin-dogness, etc.*) is the narrow content of my mental symbol 'dog'. People share this narrow concept if they have a symbol-and-mechanism pair which picks out the same set of properties in the same set of worlds. The range of mechanisms that will do so is presumably large, so identity of covariance-causing mechanisms is sufficient but not necessary for identity of narrow content. And, of course, mere syntactic identity of mental symbols is *neither* necessary *nor* sufficient. (1991a, p. 269, Reply to Block).[8]

It bears emphasis that narrow contents are specified without reference to the (non-semantic) identity of internal symbol tokens that bear them. Two people share a narrow content just in case they have a mechanism and a symbol pair (*any* mechanism and symbol pair) that effects the same mapping from worlds (contexts) to broad contents. Instantiation of the same mapping (purely extensionally defined) is then both necessary and sufficient for identity of narrow content.

This move gets rid of the problem of restricting the applicability of narrow content only to twins, i.e. only to molecular duplicates. But once the constraint of physical identity is dropped, determining whether people share the same narrow content remains problematic *in practice*. This is so despite the fact that the notion of narrow content as a mapping from contexts to broad contents seems well-defined *in theory*. I will come back to this point later on when I take up the issue of how broad content is thought to be determined.

Such was the picture Fodor had of the problem space. Twin Earth cases posed a problem for the kind of intentional psychology he envisaged, so he came up with a non-functionally specified notion of narrow content to fix it. The resulting kind of narrow content was still intentional because its identity conditions essentially depended on the broad content a mental state would have in different contexts. In other words, as Fodor put it, Twin Earth cases didn't break the connection between thought and denotation but only relativized it to a context. Everything, on the face of it, looked just fine, at least for a while.

3. *Fodor's Treatment of Frege Cases (1985–91)*

There remained really puzzling corners in Fodor's picture, however, the most puzzling of which perhaps was his treatment of Frege cases. It is usu-

---

[7] For a similar point, see Wilson 1995, pp. 237ff.

[8] Emphasis in the original. In all the quotations to follow the emphases will be original unless otherwise stated.

ally believed that Frege cases are an important test for any scheme of psychological explanation proposed to be carried out in terms of the intentional properties of mental states, since such cases involve states with the same broad content but quite different internal behaviour. Hence the usual lesson drawn is that appealing to only denotational content can't be sufficient to explain the quite different psychological behaviour of the mental states involved in Frege cases. And the usual solution has been to appeal to some notion of narrow content of such states in the explanation of their behaviour. In fact, even Fodor himself seems to think in his (1994) that one of the jobs of narrow content is to account not only for twin cases but also for Frege cases, as well as for the semantic/psychological peculiarities of so-called deferential concepts.[9]

But Fodor, after around 1985, has never appealed to *his* notion of narrow content in the explanation of Frege cases! In other words, he apparently had a different scheme of explanation for Frege cases, different than appealing to intentional laws that advert to narrow contents of the mental states they subsume. This is a real curiosity, but there were very good reasons for it: Fodor's notion of narrow content didn't work for Frege cases.

Suppose that the two Mentalese sentences #Cicero is bald# and #Tully is bald# are tokened in Smith's head. Smith doesn't know that Cicero = Tully. Then there are bound to be instances where the two tokens will behave quite differently in Smith's head and will cause Smith to do and say all sorts of things that he wouldn't if he knew that Cicero = Tully. A notion of narrow content conceived on the basis of the functional roles of these tokens can explain perfectly well why these two tokens have all these psychological differences despite the identity of their truth-conditions. But Fodor's mapping notion of narrow content can't appeal to their functional roles, nor to the physical differences between the tokens. According to Fodor's account of narrow content, all there is to evaluate whether they have the same narrow content is what mapping they instantiate from contexts to truth-conditions. But on this criterion, the two sentence tokens in Smith's head turn out to have the same narrow content, because in all contexts for which the mapping is defined they turn out to have the same truth-conditions since Tully *is* Cicero, and necessarily so.

One might think that the differences in the non-semantic type-identity of underlying sentence tokens could resolve the problem of assigning them the same narrow content. Well, those differences would indeed, let's admit, make a difference in distinguishing the two states of Smith when he thinks that Tully is bald and thinks that Cicero is bald. But this doesn't help in our desire to make it the case that they are different states *in virtue of* their having

[9] Although I will come to it below, it is interesting to note for the moment that what Fodor explicitly and officially rejects in his new book 1994 is, strictly speaking, not the notion of narrow content he previously held (i.e. a mapping from contexts to broad contents) but rather a functionalist notion of narrow content which, he says, is 'metaphysically constituted by' computational role (see especially p. 49).

different narrow content, because narrow content is externally and extensionally defined, as I was in pains to point out above. If you want the non-semantic type-identity of underlying symbols to be part of narrow content (perhaps by saying that narrow contents are functions from contexts *and* symbol types to broad contents), you lose the interpersonal ascribability of narrow content, because you have then to come up with an interpersonally ascribable identity criterion for symbol types on the basis of their non-semantic properties, which are of two sorts: narrow functional and physical. But it was in effect Fodor's despair that this can't be done for the first sort of properties that forced him to reject functional role semantics. This was, I take it, a direct consequence of his quite general resolution to reject buying anything if the price is to give interpersonally applicable identity conditions for functional roles. The remaining option was not open either, as Fodor himself acknowledges,[10] since it seems to commit one to a very strong form of type-physicalism for propositional attitudes with *specific* contents.

At any rate, nowhere in his writings during 1985–91 does Fodor actually state or even imply that Frege cases are to be handled by his notion of narrow content.[11] Instead, as I hinted above, he consistently uses another kind of apparatus: he appeals to the underlying 'syntax' of symbol tokens to keep the thoughts involved in Frege cases separate, despite the identity of their broad, as well as narrow, contents. This is most explicit in his (1989),[12] where his concern is with what Frege cases really show for the semantics of propositional attitudes. He argues, contrary to mainstream philosophy of language and mind, that the failure of substitution *salva veritate* of co-denotational expressions in opaque belief contexts doesn't show that there is more to their meaning than denotation. All Frege cases show, he claims, is that there is something else apart from denotational broad content that makes the co-denotational thoughts distinct mental states. It is left open whether this something else is semantic or not. Fodor thinks that it isn't: it is syntactic. What he has in mind here is the Mentalese symbol that is the vehicle for the broad semantic content it carries. For Fodor, the individuation of intentional mental states (propositional attitudes in particular) involves at least three elements: fixing the subject to whom the intentional state is attributed; fixing a vehicle (conceived as a Mentalese symbol token) to which the subject is said to stand in different computational relations (different relations to distinguish, for instance, believing from desiring); and fixing a broad content that the vehicle is said to bear. In Frege cases, even though the subjects and

---

[10] For the most recent statement, see Fodor 1994, pp. 107–9, although Fodor, very puzzlingly, seems to opt there for a functional specification of the type/token relation for Mentalese primitive symbols. See my paper 1996 for a criticism of Fodor on how to type Mentalese tokens.

[11] This is so despite the curious fact that in his 1994 Fodor claims that one of the jobs of narrow content is to account for Frege cases (as well as the case of deferential concepts and Twin Earth cases). More on this below.

[12] But see also Fodor 1987, pp. 86–7.

the contents are the same, the vehicles are type-distinct. Hence attributing the belief that Cicero is bald and the belief that Tully is bald is not attributing the same belief-state, despite the fact that the states attributed have identical contents. Since it is the non-semantic properties of vehicles that are supposed to be relevant for their internal computational processing, their distinct psychological behaviour and causal potential can be explained by their non-semantic (in particular, syntactic) properties. So everything seems to be fine once again.

It is noteworthy that Fodor's (1989) is one of those very rare occasions where he is very explicit that the non-semantic typing of Mentalese vehicles must *not* be carried out on the basis of their causal/functional roles:

> A vehicle is a symbol. A symbol (token) is a spatiotemporal particular which has syntactic and semantic properties and a causal role. Vehicles, like other symbols, are individuated with respect to their syntactic and semantic properties, but *not* with respect to their causal roles. In particular, two vehicle tokens are type distinct if they are syntactically different or if they express different propositions. But type-identical vehicle tokens can differ in their causal roles because the role that a token plays depends not just on which type it's a type of, but also on the rest of the world in which its tokening transpires. (This is true of the causal roles of symbols because it's true of the causal roles of everything. Roughly, your causal role depends on what you are, what the local laws are, and what else there is around.) I assume, finally, that vehicles can be type distinct but synonymous; distinct vehicles can express the same proposition. So much for the individuation of vehicles. (1989, p. 167)

If individuation by causal roles is out,[13] what is left in Frege cases is individuation by physical properties. In this passage, as in so many others, by 'syntactic' Fodor seems to mean physical.[14] As I argued elsewhere (1996), this is a real problem for Fodor because individuation by physical properties

---

[13] The rest of the article makes it clear that what Fodor has especially in mind here are the causal interactions of a symbol token *with other tokens*: 'A very rough theory of belief individuation might make do with just a person, a vehicle, and a content. You get a sharper picture if you allow in a functional role for the vehicle. Loosely speaking, I mean by the functional role of a vehicle the role that it plays in inference; more strictly speaking, I mean its causal role in (certain) mental processes' (1989, p. 168).

[14] Syntactic properties, properly understood, are properties that relate to the logical/syntactic form and the constituent structure of a symbol, e.g. the property of being a noun, or having a certain parsing tree, or being a conditional, etc. But it is obvious that these can't distinguish thoughts in Frege cases, since it is precisely the difference between particular lexical items in the same syntactic roles that causes the problem. So Fodor's appeal to 'syntax' here can't be understood in the proper sense of the term. But see his 1994, pp. 107–9. See also my paper (1997) for more discussion and for the notion of syntax the Language of Thought Hypothesis requires.

seems to commit him to a strong version of type-physicalism applicable across intentional organisms. But let's put that worry aside for the moment.

The reason why Fodor couldn't include the 'syntactic' type-identity or diversity of vehicles into his notion of narrow content was because that would explicitly commit him to a kind of content that is not interpersonally ascribable, whereas it was in fact precisely this need, namely the need to come up with a notion of an interpersonally ascribable content, that led Fodor to introduce the notion of narrow content he did in the face of the difficulties raised by twin cases. We now see that not appealing to his notion of narrow content, but to the 'syntax' of vehicles instead, doesn't help either, for precisely the same sorts of reasons.

It might be thought that all Frege cases demand for the purposes of psychological explanation is that the relevant symbol tokens be kept distinct within a single organism, i.e. *intra*personally, and this Fodor can plausibly do by appeal to the physical properties of the vehicles: some attenuated form of type-physicalism is not so implausible intrapersonally after all, and in fact, it may even be required by the Computational Theory of Mind (CTM). I want to make two quick related points against this. (It must be remembered that we are discussing the Fodor of 1985–91.)

First, we can easily construct interpersonally applicable Frege cases. Consider: Smith belives that Cicero was bald, and he hasn't heard of any Tully. On the other hand, Marvin believes that Tully was bald, and he hasn't heard of any Cicero. Also, given this difference, make them doxastically (and even neurophysiologically) similar to any degree you like. Then, given the usual background assumptions and adjustments, I ask them: 'Was Cicero bald?'. Smith says, 'Yes', and Marvin says: 'I don't know'. (If you think about it, there are indefinitely many such examples.) How is the difference in their behaviour to be explained? Not by appeal to their denotational contents, because they are the same. But since Fodor rejects appeals to vehicles' causal roles, he can only appeal to their intrinsic physical properties, which is, I take it, an implausible move that even Fodor wouldn't want to make, because it seems to commit one to type-identity theory.

Second, Fodor himself acknowledges that Frege cases generalize beyond the lexical differences created by different but co-referential proper names. In a context where he discusses almost exactly the same issue, he says:

> It should be clear that, although this problem arises for the metaphysics of Mentalese names, it can do so equally for any of the primitive, syntactically simple expressions of Mentalese on the (surely not implausible) assumption that such expressions can be coreferential. For this reason, I won't even consider proposals that depend on assuming that CICERO and TULLY are actually syntactically complex in Mentalese (e.g. that they are descriptions). Whether or not names are primitive, *some* expressions of Mentalese must be, and the present problem will arise for them, whichever they are. (1994, pp. 106–7)

I think Fodor is right about this.[15] My point is that, given the generality of this problem, there is also no reason to think that cases where we may want to appeal to non-semantic type-distinctness of underlying tokens will occur only *intra*personally. In fact, Fodor's externalism had better leave this open, because it leaves open the possibility of quite different underlying psychological mechanisms and symbols sustaining the same mind–world relations. In all these cases you will have broad content (hence narrow content) identity without necessarily the same internal functional roles, hence possibly different causal potential and different behaviour (in all the causal aspects of symbol tokens *other* than those that sustain mind–world relations). If this is right, these differences need a non-semantic explanation; thus, the same problem of individuating symbol tokens non-semantically and interpersonally is still with him.

At any rate, I want to put the discussion of this problem aside for present purposes. All we need to attend to now is that Fodor appeals to something *other* than narrow (and broad) content of intentional states in handling Frege cases.

This raises a question about the proper or canonical form of psychological explanations. In his (1987) and in so many other places, Fodor is very sanguine about intentional explanation in psychology, where, as I indicated at the beginning, this means subsuming the phenomena to be explained under intentional laws, i.e. laws that advert to the narrow content of psychological states. To put it differently, events get explained in so far as they fall under intentional laws *qua* having narrow contents whose nomological relations with each other and with stimuli and behaviour are detailed by the same laws that advert to them. But we now see that for Fodor this was only part of the story; it by no means exhausted all the available forms of psychological explanation. This is not a criticism, but just a clarification of a point buried in those of Fodor's writings (1985–91) which seem to explicitly announce his official position on psychological explanations.[16]

---

[15] Although this quote is from his recent writing (1994), I am nevertheless including it to support something he said earlier, because the claim involved equally and correctly applies to his earlier (1985–91) views about Mentalese.

[16] Fodor's implicit commitment to at least these two forms of psychological explanation all the while being explicit about only one of them, namely that psychological explanation is a matter of subsumption under psychological laws that advert to narrow contents of mental states, has created serious confusions in the literature. For instance, Stich 1983, 1991 and Devitt 1991 accuse Fodor of 'having it both ways', where this means that Fodor can't appeal to the semantic properties of thoughts in the explanation of their behaviour while simultaneously subscribing to the Computational Theory of Mind (CTM), according to which thoughts have their internal causal powers solely on the basis of their syntactic (non-semantic) properties. Fodor's response wasn't quite adequate. He appeals to the difference between the levels to buttress his attempt to have it both ways: psychological laws are all intentional, but the implementing mechanisms are syntactic/computational (see Fodor, 1987, p. 166, fn 3, and 1991a, reply to Devitt). This, however, left the difficult issue of explaining the conundrum of Oedipus quite in the dark. At a minimum, when we explain Oedipus' behaviour in the usual folk way, we don't seem to be doing implementational psychology.

On the other hand, given the difficulties mentioned above, it wasn't clear what the form and principles of the explanation of Frege cases were. Oedipus didn't want to marry his mother but he ended up marrying her all the same. He didn't want to kill his father, but he ended up killing him all the same. The common-sense explanation is quite straightforward in such cases. Oedipus didn't know that Jocasta was his mother; he didn't know that this arrogant and quarrelsome traveller was his father. I don't think that the truth of the common-sense explanation in such cases can be seriously doubted. It is therefore important that Fodor's story should be able to accommodate it. On Fodor's story, when the folk explain Oedipus' behaviour by crucially attributing to him the belief that Jocasta ≠ Oedipus' Mother, what makes this folk attribution true is, *inter alia*, the fact that the underlying Mentalese vehicles, #Jocasta# and #Oedipus' Mother#, are physically type-distinct. I happen to think that there is in fact something to this suggestion, but on the face of it, it is not at all clear how the folk manage to convey this fact. What are the underlying principles? In his (1989), Fodor suggests that the folk attribution of the belief that Jocasta ≠ Oedipus' Mother succeeds in conveying that the underlying vehicles are type-distinct by displaying them. And this in turn is done, he says, by choosing an English sentence for the embedded that-clause that is structurally isomorphic to the Mentalese one. But what ensures this isomorphism? Fodor seems to think that there is no precise answer to this question, that it is all pragmatics.[17] But then how is the scientific counterpart of the explanation supposed to go? Can we expect a scientific intentional psychology to fare any better? Well, whatever the answer was, Fodor had more work to do to clarify the kind of psychological explanation involved in Frege cases, given that it deviated from the norm he envisaged, namely subsumption under narrow-intentional laws.

As we will see in a moment, the situation is in no way different with Fodor's 'new' view that laws of psychology are broad: he has exactly the same problem. This time, however, he is explicit about what to do with it. He proposes that we ignore the problem in general. His new position involves abandoning the hope of a scientific psychology that is keyed to handling Frege cases. In other words, he seems to think that there is no future for a scientific psychology committed to *de dicto* attribution of intentional properties to organisms whose behaviour constitutes the application domain of its laws. More on this below.

This is, then, in essential outline, how Fodor thought about intentional content (narrow and wide) and its relation to psychological explanation during 1985–91.

4. Fodor's Rejection of Narrow Content

In his new work, *The Elm and the Expert* (1994), Fodor notoriously rejects the notion of narrow content of which he has been the most ardent champion

[17] See Fodor 1989, pp. 170ff.; cf. Fodor 1994, pp. 111–2.

over the years. He is, however, cautious about the manner in which he does so. His point, he says, is not to show that the notion is incoherent, or that the laws of psychology are not narrow. Rather, his aim is to show that the standard considerations for the necessity of narrow content are inconclusive, and that therefore narrow content is *perhaps* superfluous.

Yet, there are real puzzles about the way Fodor reconstructs the problem space, especially *vis-à-vis* his own earlier notion of narrow content. Here is what he does. He takes up, one by one, all the major cases that are thought to motivate narrow content and attempts to show that none of them precludes the possibility that psychological laws are *de facto* broad. The cases are all too familiar, so I won't rehearse them here in any great detail. They are cases that involve Twin Earth thought experiments, cases involving so-called deferential concepts (i.e. concepts whose semantics are said to depend on the knowledge of experts in the linguistic community), and Frege cases. Let me take the last one first.

4.1 *Frege Cases (1994)*

The real curiosity of Fodor's discussion is most apparent here. He says that Frege cases motivate the introduction of narrow content and thus arguments for the claim that psychological laws are narrow, in that they advert to such content. But the notion of narrow content that Frege cases motivate is not the one he himself developed earlier, as we saw above. It is rather a notion of content that is 'supervenient on' (1994, p. 50) or 'metaphysically constituted by' (1994, p. 49) computational role. So, strictly speaking, when he rejects narrow content he is not rejecting narrow content conceived as a partial mapping from contexts to broad contents. For instance, he writes:

> It is true by stipulation that 'Smith believes Fa' and 'a = b' doesn't entail 'Smith believes Fb' if intentional content is supposed to be narrow. Prima facie this is an advantage of the narrow construal; it allows Smith's attitude to Fa to be one thing and his attitude to Fb to be quite another . . . (1994, pp. 39–40)

But this is simply *not* true of his own mapping conception of narrow content! I spelled out the reasons why above. Here, however, it is more obvious that the specification of the function from contexts to truth-conditions can't include appeal to the type-distinctness of the underlying Mentalese sentences, since such a notion of narrow content would fail to apply to every one in the same situation as Smith. But if so, Smith's belief that *Fa* has the same narrow content as the belief that *Fb* (if he were to hold it), because they effect the same (partial) function from contexts to truth-conditions. So as far as Fodor's own views are concerned, then, nothing much has changed: he is arguing against a notion of narrow content he never held (at least after 1985).

And the way Fodor now wants to handle Frege cases is still more or less

the same as the way he wanted to handle them before, with the following twist. He argues that no intentional psychology (broad or narrow) could allow the proliferation of Frege cases because if such cases were allowed to proliferate, no intentional psychology would be in a position to explain how 'rational behavior is, generally, pretty successful as a matter of fact' (1994, p. 41); in other words, the routine success of much of daily behaviour would be sheer accident. Put differently, Fodor argues that an unchecked proliferation of Frege cases would break the connection between the rationality of an action and the likelihood of its success. His discussion is difficult to follow, and there seem to be many distinctions overlooked (e.g. matters of meaning vs. matters of truth). It is not clear, for instance, what Fodor wants to say about Oedipus' behaviour. Sometimes he gives the impression that he wants to blame Oedipus as irrational since some of his crucial beliefs were false and some of his wants were incoherent. But I think that all the ordinary intuitions about rationality go against him here.[18] Sometimes, however, he writes as if he believed that Oedipus was not irrational, but that it would be a mistake to take his case as the norm on which to construct an intentional psychology.

For whatever Fodor's discussion is worth here, the important point to observe is that it has very little to do specifically with narrow content *per se*. By Fodor's own admission, Frege cases are equally a problem for a broad intentional psychology, since, first, a broad psychology should also be in a position to explain how Frege cases are possible given that they occur, however rarely, and second, it should equally supply a mechanism to ensure that such cases don't occur very often, hence to explain why the predominance of successful action isn't an accident.

And, in fact, Fodor's attempt to account for how a broad psychology could accommodate the occurrence of Frege cases is exactly the same as his previous account. He appeals to the 'syntax' of the underlying Mentalese symbols he calls 'modes of presentation' (1994, p. 47ff.). Since the vehicles are different, Smith's belief-states (that *Fa*; that *Fb*) are different despite the identity of their truth-conditions. Nothing new here! In fact, Fodor even worries that:

> [on] the account I have been considering . . . [i]f content is broad, then behavior is only determined by content taken *together with*

[18] Rationality doesn't require that the beliefs on which we act are pretty generally true. How could it? Given that we are representational agents, all we can do is ensure that our beliefs are epistemically justified—indeed as Fodor himself seems to insist, see his 1981, pp. 241–3; 1989, p. 176, fn 10. The rest depends on the cooperation of the world. Surely, that our behaviour tends to be successful isn't accidental: rational agents are usually epistemically responsible, i.e. they try to make the best justified judgements they can given the available evidence. And, of course, epistemically justified beliefs are more likely to be true than unjustified ones. Hence, the success of our behaviour. However, this is all epistemology and has nothing in particular to do with semantics. For more discussion and criticism of Fodor on Frege cases, see Prinz 1995.

> *modes of presentation*. But if Turing was right about psychological processes being exhaustively syntax driven, it looks, once again, as though *content* per se drops out of psychological explanations. It seems that it's the syntactical properties of modes of presentation that are doing all the work, and the attachment to an *intentional*, as opposed to computational, level of psychological explanation is merely sentimental. (1994, p. 50)

The puzzle here is closely related to Fodor's ambivalence about the proper or canonical form of psychological explanations pointed out above. If behaviour is determined by content taken together with modes of presentation, how could the laws of psychology be *only* adverting to broad content? Especially given that there are *actual* cases, by Fodor's own admission, where the explanation of behaviour seems to require adverting to modes of presentation. As I indicated earlier, his answer is that the best a scientific psychology can do about Frege cases is to explain them *away*, since any appeal to differences in the underlying modes of presentation can't be principled. Hence, there is no serious hope for a scientific psychology whose laws can cover Frege cases.[19]

Let's put this aside. All I want to emphasize here is that nothing of importance has changed, as we have seen, between the Fodor of (1985–91) and the Fodor of (1994). His notion of narrow content had nothing to do with the explanation of Frege cases, and these cases still receive the same sort of treatment.

4.2 Experts

The way in which deferential concepts motivate narrow content is controversial. But the story is supposed to go like this. Although I seem to have the concepts of ELM and BEECH I can't distinguish elms from beeches, where this used to get translated as: I have no particular belief that is true of elms but not of beeches, and vice versa. My concepts are therefore 'narrowly' identical, but nevertheless their extensions are different, as they are said to be determined by the prevailing conventions of one's linguistic community according to which, e.g. the extensions of 'elm' and 'beech' are fixed by the experts who live in that community, and to whom the rest of the community defer. What we have here looks like an *intra*personal version of twin cases:

---

[19] Thus, Fodor's new position is indeed new *vis-à-vis* the methodological solipsism of his 1981 (in fact, quite opposite to his views there), where his vision of scientific psychology was shaped by the explanatory model he thought Frege cases required. There he attempted to assimilate the *de dicto* intentional content required for the explanation of Frege cases to a kind of narrow content constructed purely out of internal computational roles. Fodor's (half-conscious) realization that there can't be a robust identity criterion for interpersonally applicable computational roles is the key to understanding the development of his thought over the years. See, again, my 1996 for more elaboration and discussion of this point.

internally identical concepts in one and the same head whose extensions differ nevertheless. So, it's alleged that since the internal psychological behaviour of these two concepts is type-identical they should be subsumed by the same laws, but this is possible only on the assumption that the laws advert to narrow contents, because narrow contents are constructs out of internal behaviour (more accurately, out of functional/conceptual role).

Now, again, in arguing against the case deferential concepts are supposed to make for narrow content, Fodor seems to have forgotten that his own notion of narrow content wasn't based on functional role. His discussion is not easy to follow, but he seems to be claiming that even for the notion of a narrow content based on functional roles, the narrow contents of ELM and BEECH are in fact different since the very internal mechanisms that would cause us to defer to experts separately for each of these concepts are different. In this way, Fodor seems to neutralize the argumentative force of the case for narrow content made by deferential concepts, since on both broad and narrow conceptions ELM and BEECH come out to have different content.

As far as I can tell, Fodor nowhere discusses the case of deferential concepts *vis-à-vis* his mapping notion of narrow content.[20] But as can be easily seen, on this notion of narrow content ELM and BEECH come out to have different narrow contents, since they instantiate different functions: assuming that internal Mentalese symbols for 'elm' and 'beech' are 'syntactically' type-distinct, #elm# has either no broad content or a different one (*elm*) in every context in which #beech# has *beech* as its broad content.

So it seems that on both construals of narrow content, the case of deferential concepts could receive no different treatment by Fodor *vis-à-vis* his previous position. Nothing new here either.

4.3 Twin Earth Cases

That twin cases are and have been Fodor's main worry becomes instantly obvious once we see that Fodor does indeed seem to have changed his mind about how to treat them. I will argue, however, that the change is only apparent.

I will first take up Fodor's position in *Psychosemantics* and show that his way of handling twin cases on the basis of his mapping notion of narrow content was problematic in that this notion collapses into the notion of broad content that results from the information-theoretic account of mental representation he develops in the fourth chapter of that book. I know this sounds bizarre, but as we will see, this is indeed what happens in *Psychosemantics*. Once I make this claim stick, the rest of my argument is relatively easy. Who wants such a notion of narrow content that keeps turning into

[20] See, however, his reply to Loar (1991a, pp. 285–6), where he briefly takes up the issue of deferential concepts *vis-à-vis* his account of intentionality.

broad content? And when Fodor has decided to dispense with narrow content, it is therefore important to be clear about what notion of narrow content he was getting rid of. In *The Elm and the Expert*, when he announces that narrow content is superfluous and that perhaps we could make do with only broad content, it is also important to realize that the notion of broad content he wants to work with here is the very notion into which the rejected notion of narrow content has kept collapsing. This is more or less the sense in which I want to say that nothing much substantial has changed as far as Fodor's attitude toward intentional explanation is concerned.

Well, let's first see what the change was supposed to be. I have already said what Fodor previously thought about twin cases and why: he thought that they require a narrow intentional psychology. In his new work, he no longer thinks that. Instead, he thinks that since twin cases are nomologically impossible, a broad psychology ought not to bother with them. For, he says, 'empirical theories are responsible only to generalizations that hold in nomologically possible worlds' (1994, p. 29). Now, I will have a few words to say about this later on, but for the moment, notice that he thinks that the same laws that advert to broad content can't cover twins since twins have different broad contents. So he must be thinking that his information-theoretic account of broad content assigns distinct broad contents to twins. This is why he is making such a not-so-persuasive move by saying that a broad psychology could safely ignore twins. Once I finish my discussion of the strange interplay between his notion of narrow content and the notion of broad content that results from his information-theoretic account, we will see that there is a better way for Fodor to handle twin cases, a way for which all his intuitions have been crying deep down for years. Previously he wanted to handle twin cases *via* a notion of narrow content. Now he thinks that was unnecessary: twins are really a red herring as far as the scientific prospects of a broad intentional psychology are concerned.

So then, here is what happens in *Psychosemantics*. Narrow content as developed in the second chapter of the book is a (partial) function from contexts to broad contents. It is not semantically evaluable content since semantic evaluability apparently requires a context, and the assignment of narrow content is in a certain sense context-independent, or more accurately, independent of any *particular* context, since you get broad content only when you supply a context. There is an intuitive sense in which narrow content is the disjunctive sum-total of all the broad contents a mental state would have in each and every context for which the function is defined. As Fodor says:[21]

> 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

[21] See also Fodor (1987, p. 52).

> 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, p. 269, Reply to Block)

Strictly speaking, then, narrow content ain't in the head! For the set of ordered pairs in question is obviously not in the head. We need to make a distinction here between the narrow psychological state, namely a symbol-and-mechanism pair, and the narrow content, namely the set of <context, broad content> pairs, that gets assigned to that 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, as the individuation of narrow psychological states on the basis of their narrow content is extensionalist/externalist (see below).[22]

When Fodor talks about narrow content's being in the head (e.g. 1987, ch. 2), what he probably means to say is that the narrow state, whose individuation conditions are given extensionally/externally, is in the head.[23] The narrow internal state that becomes individuated by such a notion of narrow content is in fact an equivalence class of particular symbol-and-mechanism pairs. 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 to covary with a range of external stuff in different contexts. The point about Fodor's mapping notion of narrow content is that these internal dispositions are individuated externally. *Any* syntactic-object/mechanism pairs which effect the same mapping from contexts to broad content have the same narrow content.

However, individuation of a mental state by narrow content presupposes individuation by its broad content in a given context. So Fodor needs an independent criterion of broad content, i.e. he needs to tell a prior story about what the broad content of a certain mental state is in a particular context, and why. Since the identity of narrow content is fixed externally, on the basis of what broad content a given state would have in a particular

---

[22] I should point out that the claim here isn't that narrow content can be externally *represented*. This characterization is meant to be contrasted with Burge's characterization of individualism/internalism: 'According to individualism about the mind, the mental natures of all a person's or animal's mental states (and events) are such that there is no necessary or deep individuative relation between the individual's being in states of those kinds and the nature of the individual's physical or social environment's (1986, pp. 3–4). According to this characterization, Fodorian individuation of a narrow psychological state is *essentially* externalist: it's essentially whatever bears a narrow content, i.e. a set of ordered pairs. The identity of the set is in turn given by essential reference to elements of the pairs that are external, in just Burge's sense, to the agent.

[23] If this is right, then, contrary to what Fodor says in the second chapter of 1987, narrow content turns out to be expressible in principle after all.

context (for which the narrow content is defined), Fodor needs to say in virtue of what that state has the broad content it does in that very context.

Fodor first sketched such a theory in the fourth chapter of the very same book, *Psychosemantics*.[24] Not only did his formulations of the theory change over the years; the theory itself kept changing. In *Psychosemantics*, we have a purely counterfactual construal of the nomic covariation of the instantiation of two properties, say, the property of being a cow (*cow*) and the property of being a #cow# token in one's Mentalese. In *A Theory of Content* (1990a), Fodor briefly experimented with a clause that requires that #cow# be actually/historically causally connected to *cow*. But in his (1994) he rejected this clause altgether and returned cheerfully to the original version of the theory, to 'pure informational semantics'. I will take up the theory in this form. In fact, as we will see, it is crucial for me that Fodor's theory of broad content is a *pure* informational account of broad content.

According to the theory, a Mentalese predicate #F# expresses the property *F* (#F# has the broad content *F*) IF

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

The intuitive idea and formulation I will stick to in what follows is that the broad content of a symbol is a matter of what it *would* causally correlate with and what the objective/synchronic dependencies are among the dispositions sustaining these correlations.

Now I can say what the problem is. As the content-making correlations are stated purely counterfactually, there is a clear prima facie sense in which the theory assigns broad content to symbols *independently of particular contexts*, or equivalently, *without regard to what particular context the subject actually happens to be in*. Now there are a couple of ways in which one might try to curb the unwelcome consequences of this for twin cases, as indeed Fodor himself seems to have attempted. I will come to them below. For the moment, let me focus briefly on the notion of broad content afforded by this theory.

If this is indeed the account of broad content on the basis of which narrow content is to be determined, there is almost no point in retaining the notion of narrow content, since it seems to collapse into the notion of broad content. Or perhaps, vice versa. There is a clear sense in which broad content understood along the above lines individuates the mental states (symbol-and-

[24] Here I am ignoring Fodor's previous stabs at the problem of naturalizing intentionality. See, especially, his c. 1984.

mechanism pairs) of an organism irrespective of the actual context the organism happens to be in. The theory assigns broad content to a symbol-and-mechanism pair on the basis of the counterfactual behaviour of that pair, i.e. on the basis of what it would nomically covary with across different contexts. In other words, 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 behaviour.[25] So to individuate a mental state on the basis of its broad content is 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 this was exactly the characterization of narrow content given by Fodor earlier![26]

Furthermore, what makes the counterfactuals of this purely informational theory of broad content come true is (mostly)[27] in the head, and importantly so: it is certain mechanisms that fix the disposition to token a certain symbol

---

[25] Cf. Fodor's discussion of how God would know what the broad content of one's mental state is 1990a, pp. 125–7.

[26] More strictly, since the narrow content of a Mentalese expression, #F#, is officially specified as a set of ordered pairs, $<w,d>$, whose first element (w) is a context and the second a broad content (*de*notation) #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, *prima facie* 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 context, 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 {<Earth, H_2O>, <Twin-Earth, *XYZ*>, . . . }. But his informational semantics prima facie assigns the disjunctive broad content (H_2O *v XYZ v* . . .) to our respective #water# tokens. If so, the narrow content of our #water# tokens prima facie is {<Earth, (H_2O *v XYZ v* . . .)>, <Twin-Earth, (H_2O *v XYZ v* . . .)>, . . . }. So the narrow content of #water# as a partial function from contexts to broad contents is prima facie 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. As advertised, below I will argue that what is prima facie true of his informational semantics remains true even after Fodor's attempts to block this obvious consequence of his theory.

[27] This qualification may be thought tendentious. But what I have in mind here is Fodor's idiosyncratic 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, pp. 118–9; cf. also 1991a, pp. 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. So it is not clear at all how the extension of the theory is really supposed to go. I think this attempted extension of the theory is untenable. But I will have to leave its discussion for some other occasion.

when a certain property is suitably instantiated.[28] Put differently, instantiation of a certain property nomically controls the tokening of the symbol that expresses the property. But because this control 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.[29]

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 such as UNICORN and CENTAUR.[30] No unicorn or centaur has ever actually existed, and it is likely that none will ever exist actually. 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* instantiations of these properties.

Let me briefly 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 seems, in a non-trivial sense, to be an internalist naturalism, contrary to what Fodor himself says in his (1994).[31] In other words, what

---

[28] 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, p. 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, p. 120).

[29] 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.

[30] See Fodor 1987, pp. 163–4, fn 5 and 1994, pp. 115–19.

[31] 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, pp. 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

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, 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). So the sense of internalism involved here this time isn't that content can be specified without essential reference to contexts/environments. Rather, it is the sense that emerges when we reflect on how the theory assigns a broad content to an internal state: the grounds of assignment lie squarely within the skin of the organism as the internal dispositions counterfactually get connected to a variety of broad contents in different contexts. The contribution of the actual context isn't privileged over counterfactual ones. So the specification of content isn't done by reference to any *particular* context (actual or not).

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 resides in the head.[32] I believe that Fodor's claim (1990a, b) that meaning (content) is robust is to be understood in this latter sense, which is, I think, intimately connected to his dispositional and 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

---

way his purely informational theory handles them as an argument for it because, 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 robust sense.

[32] As I said above, 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 it 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). Similarly for 'broad content'. Although I point out the ambiguity here, in what follows I will be largely ignoring the distinction I've just drawn, leaving to the context the job of disambiguating 'content' or 'meaning'.

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).[33]

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. I think that, at least in *Psychosemantics*, Fodor really didn't realize this consequence of his theory. Take, for instance, his treatment of Putnam's 'grug' example at the end of the third chapter of that book:

> When Elmer and Oscar start out—when, intuitively speaking, they have the same beliefs about 'grug'—theirs is just a Twin case: *different wide contents* because of the difference in contexts, but *the same narrow contents* because there is the same mapping from contexts onto truth-conditions realized in each of their heads . . . as they get older, however, things change. Whereas at first tokenings of 'grug' would have been elicited from *either* child by *either* aluminum or silver, at the end only silver controls 'grug' for Elmer and only aluminum controls 'grug' for Oscar. So at the end, Oscar and Elmer are different functions from contexts to extensions, and the narrow contents of their concepts differ accordingly. (1987, pp. 93–4)

Given the pure informational theory of broad content he develops in the chapter that follows the above passage, it is simply *not* true that initially the broad content of #grug# in Elmer is *silver* and in Oscar *aluminum*. As far as the discussion in *Psychosemantics* is concerned, even though the actual contexts are different, the tokening of #grug# is controlled by *silver or aluminum*, i.e. the broad content of #grug# is disjunctive for both children since if Oscar *had been* in Elmer's context, his #grug# *would have been* controlled by silver, and vice versa. Curiously, Fodor, in *Psychosemantics*, fails to take into account the relevant counterfactuals he is otherwise so keen to point out.[34] Here we have the most paradigmatic case: initially both children have the same narrow *and* broad content, and they end up having different narrow *and* broad content! That is because, given the notions as specified by Fodor, they collapse into one. Or rather, if we stick to the purely informational account of

---

[33] For a more detailed elaboration of the relation betwen Fodor's notion of robustness and intentionality in Brentano's sense, see my (in preparation), where I lay out the foundations of an informational semantics that would accommodate both internalist and externalist intuitions.

[34] See, for instance, Fodor's criticism of Dretske's account of error in his 1984, and his reply to Baker in 1991a; Fodor 1990a, p. 58, pp. 62–3.

broad content, we may say that the notion of narrow content has just merged with the notion of broad content.[35]

If Fodor wants to resist this conclusion, he has to find a way to handle twin cases so that twins don't wind up with disjunctive broad content. Indeed, this is what he attempts to do in a few places. In what follows, I will argue that none of them succeeds.

Let's start with what he says in his (1994). His official position here *vis-à-vis* the official/original Twin Earth example seems to be that such 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. But this solution seems just too quick and cheap, and it seems that even Fodor feels the same way, because he writes:

> 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, pp. 30–1)

Now, the first thing to notice is that, given this set-up, if there are narrow generalizations that are missed (however spurious they might be—put this worry aside for the moment), that must be because there are mental states with different broad contents that nevertheless behave in twins' heads in quite similar ways that are psychologically significant. In other words, there are nomologically possible or even actual twin cases. Question: given the purely informational account, what keeps these broad contents distinct; that

---

[35] As the reader might already have gathered, I intend to be pretty much relaxed about what collapses into what. Let me repeat: the mapping notion of narrow content proposes internal psychological mechanisms (dispositionally characterized) whose individuation essentially depends on mind–world covariation in different contexts/worlds. The notion of broad content Fodor's purely informational semantics delivers also depends on mind–world covariations, and because the covariations are counterfactual, the contribution of the actual context isn't privileged over the contribution of the counterfactual ones. So, the broad content ends up being disjunctive. Furthermore, as we've seen, very much like the individuation of narrow content, the mechanisms that sustain covariations are individuated as whatever *internal dispositions* that sustain those covariations. So the (disjunctive) meaning is in the head as meaning-making dispositional mechanisms are in the head ('internalism'), but their individuation essentially adverts to a range of different (actual or possible) contexts ('externalism'). As long as these main points are intuitively clear, I think it is harmless to be relaxed about what collapses into what.

is, why do the relevant mental states not end up having the same disjunctive broad content? It seems that the situation in such cases must be very similar to Oscar/Elmer case. Perhaps Fodor wants to say: Fine, Oscar and Elmer did, in the beginning, share the same disjunctive broad content. But then broad content psychologies could subsume them under the same broad generalizations after all! They wouldn't be missed!

So it must be that Fodor has some other story to tell about why in such cases twins have distinct broad contents despite sharing all their dispositions to token concepts in much the same sorts of circumstances. And it appears that he does in fact have more than one story to tell. I say 'it appears' because Fodor's discussion is not easy to follow and, at times, puzzling from both an exegetical and theoretical point of view. So I'd like to ask the reader to bear with me in my attempt to understand, and then critically reconstruct, why Fodor thinks that twins have non-disjunctive content. I will show that at the end Fodor's attempted answers don't work.

Given the quotation above, I take it that Fodor now concedes that twin cases can occur in worlds nomologically accessible from ours. I will therefore restrict my argument to such cases and forget about nomologically impossible ones as Fodor himself does. I will consider two such twin cases.[36]

Case 1: Here is the set-up. The historical situation is pre-modern chemistry, 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 chemistry laboratories. There are two communities living in different parts of the wood. The relations between these two communities are non-existent because different parts of the wood 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 wood. 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 Mentalese tokens of #water# that differ in broad content: Smith's #water# means H_2O and twin-Smith's *XYZ*, despite the fact that they seem to share all their internal dispositions.

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

As far as I can tell, Fodor has tried three different answers to this question in different places whose mutual coherence is moot.

Fodor's Answer #1: Let's start with the answer he gives in a footnote in his (1990a):

---

[36] In what follows, as Fodor himself does in his 1994, I will continue to use XYZ in my examples for convenience, and thus assume it to be nomologically possible.

> 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 would) 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, p. 133, fn 16)

The first thing to notice here 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 admit), there can't be (really can't, not just nomologically possibly can't) any instantiations of them, whatever that means exactly. 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 H_2O 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.

There is a further difficulty. Let's grant that the #water# tokens of Smith and twin-Smith have different broad contents. So, different broad content generalizations subsume them. Nobody will have the slightest idea that Smith and twin-Smtih constitute different psychological kinds as far as the causal potential of their #water# tokens is concerned. Now, imagine that (because of, say, advancements in their technology or taboos, or whatever) they start travelling into each other's contexts; then their broad contents will become disjunctive without anyone's having the slightest idea that the laws on the basis of which they have been explaining and predicting their fellows' water/twin-water-directed behaviour have changed. And, then again, imagine that only a thousand years later they discover that the stuff in their respective puddles is different. They will say: 'Aha! All the generalizations we used to explain and predict our "water"-behaviour were wrong.' Or will they?

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. Period! Because learning what something really is changes your dispositions.[37] You cease to be the same function from contexts to broad contents, to put it in the old jargon of narrow content.[38]

---

[37] Cf. Fodor (1987, pp. 94–5) and (1991a, pp. 302–4, Reply to Stalnaker).

[38] Fodor has indicated (in conversation) that as a response to my criticism he wants to pursue the line of thought summarized here under 'Fodor's Answer #1.' If I understand him correctly (about which I am not sure), he wants to claim that among the counterfactuals that support the law 'water → #water#', only some are relevant to the truth of the law. Others are to be discarded. He doesn't 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

Fodor's Answer #2: Fodor's second answer is perhaps the most cryptic. He doesn't elaborate on it much, and sometimes what he says in adjacent paragraphs seems to conflict. The place I have in mind is (1994, pp. 30–3), where he tries to explain why nomologically possible twin cases don't constitute an embarrassment for a broad content psychology.

Here is what seems to be the crux of his answer:

> According to informational semantics, if it's *necessary* that a creature can't distinguish Xs from Ys, it follows that the creature can't have a concept that applies to Xs but not to Ys. Since informational semantics always assigns disjunctive contents in such cases, it never permits twins to arise in respect of them. Informational semantics does, however, allow you to have a concept of Xs even though you don't distinguish Xs from Ys *when your failure to distinguish them is accidental*; i.e. when there's no law that says you can't distinguish them. If, in such a case, you have a twin who also accidentally fails to distinguish Xs from Ys, but whose concepts applies to Ys and not to Xs, then a content theory would miss a psychological generalization that subsumes the two of you, *viz.*, that you both apply C to both Xs and Ys. But missing this generalization argues in favor of the theory since, by assumption, it's accidental that the generalization holds. (1994, pp. 32–3)

That is, #water# tokens of Smith and twin-Smith have different broad contents *because* although they don't distinguish H_2O and XYZ samples, in other words, although they would apply their #water# to either of them (i.e. they have the same dispositions), it is accidental that this is so, i.e. it is not nomologically necessary that they don't distinguish them. So if their contents were disjunctive, a broad content psychology would capture them; but since their broad contents are different, their behaving similarly will be missed by a broad content theory that assigns non-disjunctive contents to them. But Fodor is happy with this because, oddly enough, he thinks that this consequence makes his theory better.

Part of the problem with this answer is that saying that it is accidental that they fail to distinguish H_2O and XYZ is not the same thing as saying that it's not nomologically necessary that they do so. To see this let's change Case-1 in the following way.

Case 2: Stipulate that Community *B* has both H_2O and XYZ around, and let everything else be just the same as before.

---

in need of a notion of local law that is committed to making this distinction among counterfactuals. I am not sure that this is correct, but in the absence of a more elaborate suggestion, I am inclined to leave my criticism above as it is. See, however, my (in preparation) for both an elaboration of this kind of reply and the potential difficulties associated with it in the light of a broader discussion of the notion of information.

Now if the sense in which it is accidental that, in the previous (Case-1) scenario, Smith and twin-Smith don't distinguish H_2O and XYZ, and *therefore* have different, non-disjunctive, broad content, is to be granted to Fodor as intuitively plausible, then it should be more plausible to say that because it is equally accidental (in fact, even more so, whatever exactly that means) that the members of Community *B* don't distinguish H_2O and XYZ, their #water# tokens must be non-disjunctive! But this doesn't make sense. Their #water# tokens, by Fodor's own admission, have disjunctive content *H_2O or XYZ*. Here is what Fodor says in a passage adjacent to the quotation above:

> Suppose, for instance, there *is* XYZ after all, not just on Twin-Earth but also around here. So the local puddles are sometimes XYZ and sometimes H_2O, depending on which it happens to have been raining. And let it be that, just by sheer luck, all the puddles that you have encountered . . . have been puddles of H_2O. Intuition suggests (anyhow, mine does) that in this case C [#water#] is not a kind concept for you, *viz.* that XYZ and H_2O are both in the extension of your C-beliefs. This is as it should be since, presumably, the One True [Purely] Informational Semantics will dictate that the dispositions in virtue of which your concept C applies to Xs involve *lawful* (*viz. non*accidental) regularities in your causal interactions with Xs. Correspondingly, a broad psychology will fail to capture the generalizations under which your behavior falls together with the behavior of bona fide water-believers. But this is as it should be too; it is, by assumption, just an accident that you behave like a water-believer (*viz.*, that you apply C to H_2O and not to XYZ). *Failures to capture accidental generalizations don't impugn theories.* (1994, p. 32)

This scenario is almost like Case-2. It is clear that, given this passage, Fodor would *a fortiori* assign the disjunctive content, *H_2O or XYZ*, to twin-Smith's #water# tokens. So, according to Fodor's answer under consideration, what are the principles on the basis of which the One True (Purely) Informational Semantics assigns a disjunctive content, *H_2O or XYZ*, to twin-Smith's #water#, but a non-disjunctive content, *H_2O*, to Smith's? If we stick to our original characterization of Fodor's answer suggested by the first quotation, it seems that it is because in the latter case, it is accidental that Smith doesn't distinguish H_2O from XYZ, whereas in the former case, it is no accident that twin-Smith regularly fails to distinguish H_2O from XYZ.

Fodor's suggested answer doesn't work, because it treats parallel cases differently where it ought not to. If the sense of 'accidental' is kept constant, then it is at least equally accidental in Case-2 that Smith and twin-Smith don't distinguish H_2O and XYZ samples. But if 'being accidental' is the crux of his answer, Fodor must be able to treat Smith and twin-Smith in parallel ways; yet it doesn't make sense to suggest that twin-Smith's #water# tokens have non-disjunctive content: What is the non-disjunctive broad content of twin-Smith's #water# tokens if it is not disjunctive?

Further, it is intuitively clear that it can't be accidental that they fail to distinguish H_2O and XYZ samples if, say, they will have the means to do so in a thousand years, even if the travelling problems are solved! But if it isn't accidental, then the generalizations missed can't be spurious, because it is not accidental that the twins behave in the same way! If Fodor stuck to his purely informational theory of broad content, he could tell a better and more gratifying story (from his own perspective): they behave in the same way because they have the same broad content, a disjunctive one, *H_2O or XYZ* (because their dispositions to token their #water# symbol match) until they discover the true essence of their respective stuffs. For learning what something really is changes your dispositions; you cease to be the same function from contexts to broad contents, to put it in the old jargon of narrow content!

There is a further problem with this second answer of Fodor which I will take up in discussing the third.

Fodor's Answer #3: 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 H_2O) 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 H_2O: there's a break in the XYZ/'water' connection *without* a break in the H_2O/'water' connection in nearby world where H_2O is distinguishable from XYZ. (1990a, p. 116)

I don't think this answer works. Here is why. First 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 in fact seems to be characterizing what makes our use of *linguistic* symbol 'water' mean what it does. Our concern, however, is not with the semantics of linguistic symbols but rather 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:[39]

[39] See also Fodor's reply to Loar, where he is explicit that '[t]he application of mental representations can't be deferential in this sense for at least two reasons: first, we have no policies with respect to our concepts (only with respect to our words); and, second, we think in a *de facto* private language, a policy of deference to other speakers of which would verge on incoherence' (1991a, pp. 285–6).

> 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, p. 134, fn 25)

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. But if so, 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 H_2O despite the fact that he would apply it to H_2O?

Fodor's answer must be that twin-Jones' present disposition to apply #water# to H_2O 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 H_2O where the two substances are distinguishable, i.e. there are worlds in which his #water# symbol tracks XYZ but not H_2O, and there are also worlds in which he applies #water# to both substances where they are indistinguishable.[40]

But these worlds are irrelevant to answering the above question: namely, what makes twin-Jones' *present* disposition to apply #water# to H_2O asymmetrically dependent on his *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

[40] Cf. Fodor (1990a, pp. 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.

twin-Jones' present dispositions are kept the same. Remember that the asymmetric dependences 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.[41] Similarly, how your present dispositions *will* change/evolve is irrelevant to determining the objective dependences 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 H_2O are worlds in which H_2O 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 H_2O 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. In the worlds where twin-Jones can tell XYZ and H_2O apart he has a significantly different set of dispositions than the one he has in worlds where he can't. And these differences will make twin-Jones no longer a twin of Jones, namely, 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 H_2O 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 H_2O? 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 considering the *synchronic* dependences 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 (as implicitly acknowledged by Fodor himself): 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 *H_2O or* XYZ!

In fact, this conclusion shouldn't come as a surprise. For Fodor's asymmetric dependence clause was initially introduced to explain how error (and robustness) is possible. Perhaps, it can explain how it is possible for my #water# symbol to mean *water* but not *vodka* (or, for that matter, for #horse# to mean *horse* but not *cow-on-a-dark-night*), or even the non-labelling tokenings of my #water# (cf. Fodor, 1990b). But surely it shouldn't be supposed that it can handle the *echt* twin cases where you indeed seem to have your

[41] Cf. Fodor (1987, p. 109) and (1991a, Reply to Baker).

relevant dispositions symmetrically dependent on one another, hence have a disjunctive content.

We can also see what was really wrong with Fodor's previous answer. It went something like this: as long as twins' failure to distinguish between H_2O and XYZ is accidental (not nomologically necessary), they have non-disjunctive content. Since it is nomologically possible that twins *will* eventually come to discover the true chemical nature of the respective substances, there is, if you will, a certain sense in which it is accidental that they don't actually distinguish them *now* (say, a thousand years before their scientific breakthrough, or whatever). But this is irrelevant to evaluating what objective/synchronic dependences exist among their *present* dispositions to token #water#.

5. *Conclusion*

Is there any other avenue left for Fodor to explain why twins don't have disjunctive content? Well, perhaps there is, although I don't see one. My point at this stage is that trying to come up with such an explanation is not worth the effort, for basically two sorts of reasons. First, as Fodor himself keeps emphasizing, it is built into purely counterfactually stated informational semantics that it assigns the same dispositions to physically type-identical systems, and trying to find ways for the theory to do otherwise is unnatural and seems to breed monsters, as we have seen.

Second and more importantly, why should Fodor keep trying to come up with such an explanation, given that there is a better way to accommodate the intuitions driving him to reject his notion of narrow content? In fact, given the theoretical situation he has gotten himself into, all the intuitions cry for assigning a disjunctive content to twins. This is the sense in which there isn't much of a difference between his notions of narrow and broad content. Or rather, narrow content merges with broad content.

Let us suppose, then—just suppose—that twins do indeed have disjunctive broad contents in all the relevant situations. How much trouble would that save Fodor? Quite a lot! He would, for instance, be in a position to say confidently that laws of psychology are all broad (just as his 'new' position maintains) and that we don't therefore lose any explanatory leverage *vis-à-vis* going purely narrow (so that way he wouldn't have to avoid 'narrow' explanations of twin cases). He could, for instance, withdraw his extremely unpersuasive claim that a scientific psychology ought not to bother with nomologically impossible twins.

Furthermore, given the otherwise strong intuition that meaning is something that is in the mind, the intuition that twins have disjunctive content may not really be weak. In fact, one could plausibly argue that Putnam's thought experiment doesn't really show that meaning ain't in the head; what it shows is that the meaning that is in the head is disjunctive. I know that such a claim is extremely controversial. I am not saying that it's right—

although I am not sure that it's wrong either. My point is that this was the very intuition that has driven Fodor to claim that what Putnam's experiment really shows is not that meaning as such ain't in the head, but rather that *broad content*, which is to be fixed by both the (narrow) meaning in the head and a context, ain't in the head.[42] What we have here, it appears, is just the same intuition; for there is an important sense in which Fodor's purely informational theory of broad content makes having a broad content a matter of having meaning in the head, since it is a purely dispositional theory, i.e. it assigns broad content purely on the basis of what dispositions one has. And having the relevant dispositions is a matter of having things in the head. This fits perfectly well with Fodor's long-standing intuitions that meaning is in the mind and that psychology, having essentially to do with mind, is an intentional science through and through.[43]

Moreover, Fodor would perhaps be pleased to be able to reason as follows: it may not really be so crazy to say that twins have disjunctive content after all, given that twin cases, for all intents and purposes, are *extremely* rare occurrences (not to mention the nomologically impossible ones); in other words, saying that twins have disjunctive content doesn't really jeopardize the respectability of one's semantic theory if that theory otherwise succeeds in assigning the right contents *in the normal course of events*. This move would in fact follow Fodor's own otherwise quite solid intuition that twins are not in the normal course of events. So the real worry Fodor would like us to see is whether his informational theory does otherwise assign the right content in the normal course of events, which is another and quite separate issue. As such, my proposal that Fodor had better assign disjunctive content to twins (as his purely informational theory requires anyhow) puts the burden where it in fact belongs—just as Fodor does feel, or ought to feel at any rate.

So, to recap. Fodor's motivation for developing a notion of narrow content was to accommodate twin cases. And as we saw, his notion of narrow content was only responsive to the problem twin cases were thought to have created for an intentional psychology. It failed to handle Frege cases. It failed to handle the case of deferential concepts. But it wasn't really designed for them in the first place. When Fodor, therefore, has rejected his notion of narrow content and at the same time (re)adopted a *purely* informational semantics, nothing much has happened, except that he has missed what seems

---

[42] See Fodor's discussion in 1987, pp. 45–53.

[43] As indeed epitomized by the title of a recent anthology on Fodor's work: *Meaning in Mind* edited by Loewer and Rey 1991. It is important to note in this context that one bothersome consequence of a purely dispositional theory of broad content, as noted by many people, is that it puts broad content, intuitively, at the mercy of what beliefs/theories you happen to hold at a given moment—a consequence to which externalist theories were generally thought to be immune and which, paradoxically, drove Fodor away from functionalism in semantics towards an allegedly purely 'externalist'/denotational semantics! This consequence is even acknowledged, by and large, by Fodor; see especially 1991a, pp. 302–4, Reply to Stalnaker.

a better way of accommodating his own intuitions when he felt that his notion of narrow content was after all doing no work and hence unnecessary. The reason why he missed this, I think, is that he has somehow blindly stuck to the intuition that twins *must* be assigned non-disjunctive content come what may.

At any rate, as I hope to have shown, his new position isn't—'extensionally',[44] so to speak—much different in substance from his previous one, since, despite Fodor's desire to the contrary, his purely informational theory of broad content does, as a matter of fact, assign disjunctive content to twins, and as such, is in a position to do all the important theoretical work that his own previous notion of narrow content was supposed to do. And all the rest—Frege cases, deferential concepts—are still handled by Fodor in just the same sort of way as before.

So, although Fodor meant to change his mind on narrow content and thought he had done so, i.e. although 'intentionally', so to speak, there was a change in his attitude towards narrow content, 'extensionally' his new position turns out to be substantially the same!

Department of Philosophy
The University of Chicago

References

Aydede, M. 1996: Typing Mentalese Tokens: Fodor's Dilemma. Draft, The University of Chicago.

Aydede, M. 1997: Language of Thought: The Connectionist Contribution. *Minds and Machines*, 7: 1, 57–101.

---

[44] This claim, namely that Fodor's mapping notion of narrow content and the notion of broad content his purely informational semantics delivers are extensionally equivalent, should be qualified. For strictly speaking they are not: along with Fodor we've stipulated that the domain of purely informational semantics be restricted to worlds that are nomologically accessible from our world. So Fodor might maintain that whereas his previous notion of narrow content applies to nomologically impossible cases as well, the disjunctive content his informational semantics yields in twin cases is restricted to nomologically possible worlds. Although Fodor can legitimately take this line, it doesn't hold much theoretical interest as far as the basic motivations underlying the introduction of narrow content are concerned. For I took his claim that 'empirical theories are responsible only to generalizations that hold in nomologically possible worlds' (1994, p. 29) to be orthogonal to the distinction between narrow and broad generalizations, i.e. his claim, if true, holds not only for broad generalizations, but for narrow generalizations as well. But Fodor means to reject narrow content even when narrow generalizations are restricted to nomologically possible worlds (as is clear in his discussion of Earth-bound twins—see above). So I assumed in my discussion that, even for polemical purposes, restricting the worlds to only nomologically accessible ones is common ground between Fodor and myself. Thanks to Jesse Prinz and an anonymous referee for pointing this out to me.

Aydede, M. In preparation: *From Information to Intentionality.* The University of Chicago.

Brentano, F. 1874: *Psychologie vom Empirischen Standpunkt* (Psychology from an Empirical Standpoint). Leipzig.

Burge, T. 1986: Individualism and Psychology. *Philosophical Review,* 95: 1, 3–45.

Devitt, M. 1991: Why Fodor Can't Have It Both Ways. In B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and his Critics.* Oxford: Blackwell.

Fodor, J.A. 1981: Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. In J.A. Fodor, *RePresentations: Philosophical Essays on the Foundations of Cognitive Science.* Cambridge, MA.: MIT Press. (Originally appeared in *Behavioral and Brain Sciences,* 3: 1, 1980.)

Fodor, J.A. 1982: Cognitive Science and the Twin-Earth Problem. *Notre Dame Journal of Formal Logic,* 23: 2, 98–118.

Fodor, J.A. *circa* 1984: Psychosemantics, or: Where Do Truth Conditions Come From? In W. Lycan (ed.) *Mind and Cognition.* Oxford: Blackwell.

Fodor, J.A. 1984: Semantics, Wisconsin Style. In J.A. Fodor, *A Theory of Content and Other Essays.* Cambridge, MA.: MIT Press. (Originally appeared in *Synthese* 59, 231–250, 1984.)

Fodor, J.A. 1985: Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum. *Mind,* 94, 76–100.

Fodor, J.A. 1986: Banish DisContent. In J. Butterfield (ed.) *Language, Mind, and Logic.* Cambridge University Press. (Also in W. Lycan (ed.) *Mind and Cognition.* Oxford: Blackwell.)

Fodor, J.A. 1987: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge, MA.: MIT Press.

Fodor, J.A. 1989: Substitution Arguments and the Individuation of Belief. In J.A. Fodor, *A Theory of Content and Other Essays.* Cambridge, MA.: MIT Press, 1990. (Originally appeared in G. Boolos (ed.) *Method, Reason and Language.* Cambridge University Press, 1989.)

Fodor, J.A. 1990a: A Theory of Content, II. In J.A. Fodor, *A Theory of Content and Other Essays.* Cambridge, MA.: MIT Press.

Fodor, J.A. 1990b: Information and Representation. In P.P. Hanson (ed.), *Information, Language, and Cognition.* Vancouver, Canada: University of British Columbia Press, 1990.

Fodor, J.A. 1991a: Replies (ch. 15). In B. Loewer and G. Rey (eds) *Meaning in Mind: Fodor and His Critics.* Oxford: Blackwell.

Fodor, J.A. 1991b: A Modal Argument for Narrow Content. *Journal of Philosophy,* 88, 5–26.

Fodor, J.A. 1994: *The Elm and the Expert: Mentalese and Its Semantics.* Cambridge, MA.: MIT Press.

Fodor, J.A. Unpublished: Narrow Content and Meaning Holism. Draft, *c.* 1984.

Kaplan, D. 1989: Demonstratives. In J. Almog, J. Perry and H. Wettstein (eds), *Themes from Kaplan.* Oxford University Press.

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

Prinz, J.J. 1995: Is Narrow Content Superfluous? Draft, The University of Chicago.

Stich, S. 1983: *From Folk Psychology to Cognitive Science: The Case Against Belief.* Cambridge, MA.: MIT Press.

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

White, S.L. 1982: Partial Character and the Language of Thought. *Pacific Philosophical Quarterly*, 63, 347–65.
Wilson, R.A. 1995: *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind*. Cambridge University Press.