INDIVIDUALISM AND SUPERVENIENCE

Jerry Fodor and Martin Davies

I—Jerry Fodor

> After the Beardsley exhibit at the V&A, walking along that endless tunnel to South Kensington Station, I thought, why this is 'behavior'—and I had said, perhaps even written: 'where does "behavior" begin and end'?
>
> Barbara Pym

I beg your indulgence. I am about to tell you two stories that you've very probably heard before. But I don't propose to tell you why I am telling you these stories; and I don't propose to tell you the punchlines. Having once told you the stories. I will then spend most of this paper trying to puzzle out what, if anything, they have to do either with commonsense belief/desire explanation or with the Representationalist Theory of Mind (RTM). The conclusion will be: not much. That may sound pretty dreary, but I've been to cocktail parties that were worse; and, there's a sort of excuse in the following consideration: the two stories I'm about to tell you have been at the center of a great deal of recent philosophical discussion. Indeed, contrary to the conclusion that I am driving towards, it is widely held that one or both stories have morals that tend to undermine the notion of content and thereby raise problems for propositional attitude based theories of mind.

Since these stories are so well-known, I shall tell them in abbreviated form, entirely omitting the bits about the shaggy dog.

The Putnam story.

Is there *anyone* who hasn't heard? There's this place, you see, that's just like here except that they've got XYZ where we've got H_2O. (XYZ is indistinguishable from H_2O by any casual test, though of course one could tell them apart in the chemical laboratory.) Now, in this place where they have XYZ, there's someone who's just like me down to and including his neurological microstructure. Call this guy 'Twin-Me'. The

intuition we're invited to share is that, in virtue of the chemical facts and in spite of the neurological ones, the form of words 'water is wet' means something different in his mouth than it does in mine. And, similarly, the content of the thought that Twin-Me has when he thinks (*in re* XYZ, as one might say) that water is wet is *different* from the content of the thought that I have when I think that water is wet *in re* H_2O. Indeed, the intuition we're invited to share is that, strictly speaking, Twin-Me can't have the thought that water is wet *at all.*

The Burge story.

The English word 'brisket', according to the Funk & Wagnall's *Standard Desk Dictionary* and other usually reliable authorities, means 'the breast of an animal, esp. of one used as food' (from the Old French 'bruschet', in case you were wondering). Imagine a guy—call him Oscar—who speaks English all right but who suffers from a ghastly misapprehension; viz. Oscar believes that only *certain* food animals—only beef, say—have brisket; pork, according to Oscar's mistaken world view, is *ipso facto* brisketless.

First intuition: Oscar, despite his misapprehension, can perfectly well have brisket beliefs, brisket desires, brisket fears, brisket doubts, brisket qualms; and so forth. In general: If the butcher can bear attitude A towards the proposition that brisket is F, so too can Oscar. (Of course Oscar differs from the butcher—and other speakers of the prestige dialect—in that much of what Oscar believes about brisket is false. The point, however, is that Oscar's false belief that pork isn't brisket is nevertheless a brisket-belief; it *is brisket* that Oscar believes that pork brisket isn't (if you see what I mean). From which it follows that Oscar 'has the concept' *brisket* (whatever exactly that amounts to).

Now imagine an Oscar-Twin; Oscar2 is molecularly identical to Oscar but lives in a language community (and talks a language) which differs from English in the following way. In that language the phonetic form 'brisket' does apply only to breast of beef (so, whereas what Oscar believes about brisket is *false*, what Oscar2 believes about brisket2 is *true*).

Second intuition: Oscar2 doesn't have brisket-attitudes; it would be wrong for *us*—us speakers of English, that is—to say of Oscar2 that his wants, beliefs, yearnings or whatever are ever

directed towards a proposition of the form: '. . . brisket. . .'. For Oscar2, unlike his molecularly identical twin Oscar, *doesn't* have the concept *brisket*; he has the concept *brisket2* (= brisket of beef, as *we* would say).

So much for the stories. Now for the ground-rules: Some philosophers are inclined to claim about the Putnam story that Twin-Me actually *is* just like Me; that it's *wrong* to think that Twin-Me hasn't got the concept *water*. Analogously, some philosophers are inclined to say that Oscar actually *is* just like Oscar2; that it's *wrong* to think that Oscar has the concept *brisket*. (Indeed, if your theory of language is at all 'criteriological' you quite likely won't be prepared to have the intuitions that Putnam and Burge want you to have. Criteriological theories of language aren't fashionable at present, but I've noticed that fashions tend to change.) Anyhow, for purposes of discussion I propose simply to grant the intuitions. If they're real and reliable, they're *worth* discussing; and if they're not, there's no great harm done.

Second, I will assume that the Burge story shows that whatever exactly the moral of the Putnam story is, it isn't specific to terms (/concepts) that denote 'natural kinds'. In fact, I'll assume that the Burge story shows that if the Putnam story raises *any* problems for the notion of content, then the problems that it raises are completely general and affect all content bearing mental states.

Third, I take it that what's at issue in the Putnam and Burge stories is clearly *something* about how propositional attitudes are individuated; and that the intuitions Putnam and Burge appeal to suggest that the attitudes are in some sense individuated with respect to their *relational* properties. (Thus, what's supposed to account for the difference in content between my belief and my Twin's is the chemical composition of the stuff *in our respective environments*; and what's supposed to account for the difference in content between Oscar's attitudes and Oscar2's is what the form of words 'is brisket' applies to *in their respective language communities*.) So I shall talk in the following way: standards of individuation according to which my beliefs differ in content from my Twin's (and Oscar's differ from Oscar2's) I'll call 'relational'. Conversely, if attitudes are individuated in such fashion that my beliefs and my Twin's are *identical* in content,

then I'll say that the operative standards are '*non*relational'. It's going to turn out, however, that this terminology is a little coarse and that relational individuation *per se* isn't really the heart of the issue. So, when more precision is wanted, I'll borrow a term from Burge; standards of individuation according to which my Twin and I are in the same mental state are '*individualistic*'.

OK, now: What do the Burge and Putnam stories show about the attitudes?

Supervenience

Here's a plausible answer: at a minimum they show that propositional attitudes, as commonsense understands them, don't supervene on brainstates. To put it roughly: States of type X supervene on states on type Y iff there is no difference among X states without a corresponding difference among Y states. So, in particular, the psychological states of organisms supervene on their brain states iff their brains differ whenever their minds differ. Now, the point about Me and Twin-Me (and about Oscar and Oscar2) is that although we have different propositional attitudes, our brains are identical molecule-for-molecule; so it looks like it just *follows* that our attitudes don't supervene upon our brain states. Since, however, it's arguable that any scientifically respectable notion of psychological state should respect supervenience, the moral would appear to be that you can't make respectable science out of the attitudes.

I'm actually rather sympathetic to this line of thought; I think there *is* an issue about supervenience and that it does come out that we need, when doing psychology, different identity conditions for mental states than those that commonsense prefers. This doesn't bother me much because (a) redrawing these boundaries doesn't jeopardize the major claim on which the vindication of the attitudes as explanatory constructs depends viz. that scientific psychological explanation, like commonsense belief/desire explanation, is committed to states to which semantic and causal properties are simultaneously ascribable; and (b) I think it's quite easy to see how the required principles of individuation should be formulated.

All that will take some going into. For starters, however, there's this: it needs to be argued that there *is* any problem about supervenience to be solved. Contrary to first impressions, that

doesn't just fall out of the Burge and Putnam stories. Here's why: to get a violation of supervenience, you need not just the relational individuation of mental states; you also need *the nonrelational individuation of brain states*. And the Twin examples imply only the former.

To put the same point minutely differently, my brain states are type-identical to my Twin's only if you assume that relational properties like, for example, the property of *being a brain that lives in a body that lives in a world where there is XYZ rather than H_2O in the puddles* do *not* count for the individuation of brain states. But why should we assume that? And, of course, if we *don't* assume it, then it's just not true that my Twin and I (or, *mutatis mutandis*, Oscars 1 and 2) are in identical brain states; and it's therefore not true that they offer counterexamples to the supervenience of the attitudes.

('Fiddlesticks! For if brainstates are individuated relationally, then they will themselves fail to suprevene on states at the next level down; on molecular states as it might be.'

'Fiddlesticks back again! You beg the question by assuming that *molecular* states are nonrelationally individuated. Why shouldn't it be relational individuation all the way down?')

You will be pleased to hear that I am not endorsing this way out of the supervenience problem. On the contrary, I hope the suggestion that brain states should be relationally individuated strikes you as plain *silly*. Why, then, did I suggest it?

Well, the standard picture in the recent philosophical literature on cognitive science is the one that I outlined above: The Burge and Putnam stories show that the commonsense way of individuating the attitudes violates supervenience; by contrast, the psychologist individuates the attitudes nonrelationally ('narrowly', as one sometimes says) thereby preserving supervenience but at the cost of requiring an individualistic (/nonrelational/narrow) notion of content. Philosophers are then free to disagree about whether such a narrow notion of content actually can be constructed. Which they do. Vehemently.

This standard understanding of the difference between the way that commonsense construes the attitudes and the way that psychology does is summarized in Table I.

COMMONSENSE TAXONOMY (Pattern A):	PSYCHOLOGICAL TAXONOMY (Pattern B):
1 Individuates the attitudes 'relationally'; hence assumes a nonindividualistic notion of content.	1 Individuates the attitudes NONrelationally; hence assumes a 'narrow' notion of content.
2 Distinguishes: —my beliefs from my Twin's —Oscar's beliefs from Oscar2's.	2 Identifies: —my beliefs with my Twin's —Oscar's beliefs with Oscar2's.
3 Individuates brainstates NONrelationally; therefore:	3 Individuates brainstates NONrelationally; therefore:
4 Violates supervenience.	4 Preserves supervenience.

TABLE ONE: How commonsense and Cognitive Science individuate mental states (according to the standard philosophical reading.)

However, one can imagine quite a different reaction to the Twin examples. According to this revisionist account, psychology taxonomizes the attitudes in precisely the same way that commonsense does: Both follow pattern A; *both* assume principles of individuation that violate supervenience. And so much the worse for supervenience. This, if I understand him right, is the line that Burge himself takes;[1] in any event, it's a line that merits close consideration. If psychology individuates the attitudes relationally, then it is no more in need of a narrow notion of content than commonsense is. It would save a lot of nuisance if this were true, since we would not then have the bother of cooking up some narrow notion of content for psychologists to play with. It would also disarm philosophers who argue that cognitive science is in trouble because it needs a notion of narrow content *and can't have one*, the very idea of narrow content being somehow incoherent.

Alas, there is always as much bother as possible; the revisionist reading cannot be sustained. It turns out that the considerations which militate for the nonrelational individuation of mental states in psychology (and hence for preserving supervenience at the cost of violating the commonsense taxonomy) are no different from—hence no less persuasive than—the ones that

[1] Notice that taking this line wouldn't commit Burge to a violation of *physicalism*; the difference between the attitudes of Twins and Oscars supervenes on the (*inter alia*, physical) differences between their worlds. Or rather, it does assuming that the relevant differences between the linguistic practices in Oscar's speech community and Oscar2's are physicalistically specifiable (I owe this caveat to James Higginbottham.)

militate for the nonrelational individuation of brain states. This becomes evident as soon as the sources of our commitment to the latter are made clear. All this takes some proving; I propose to proceed as follows:

First, we'll consider why we think that *brain states* should be individuated nonrelationally. This involves developing a sort of metaphysical argument that individuation in science is *always individualistic.* It follows, of course, that the constructs of psychology must be individualistic too, and we'll pause to consider how the contrary opinion could ever have become prevalent. (It's here that the distinction between 'nonrelational' and 'individualistic' individuation is going to have some bite.) We will than be back exactly where we started: Commonsense postulates a relational taxonomy for the attitudes; psychology postulates states that have content but are individualistic; so the question arises what notion of content survives this shift in criteria of individuation. It will turn out—contrary to much recent advertisement—that this question is not really very hard to answer. The discussion will therefore close on an uncharacteristic note of optimism: The prospects for a psychology of content are, in any event, no worse now than they were before the discovery of XYZ; and brisket is a red herring.

Causal powers

I have before me this gen-u-ine United States ten cent piece. It has precisely two stable configurations; call them 'heads' and 'tails'. (I ignore dimes that stand on their edges; no theory is perfect.) What, in a time of permanent inflation, will this dime buy for me? Nothing less than control over the state of every physical particle in the universe.

I define 'is an H-particle at t' so that it's satisfied by a particle at t iff my dime is heads up at t. Correspondingly, I define 'is a T-particle at t' so that it's satisfied by a particle at t iff my dime is tails up at t. I now bring it about that every particle in the universe is an H-particle . . . thus! And now I change every particle in the universe into a T-particle . . . thus! And back again . . . thus! (Notice that by defining H and T predicates over objects at an appropriately higher level, I can obtain corresponding control over the state of every *brain* in the universe, changing H-brainstates into T-brainstates and back again just as the fancy

takes me.) With great power comes great responsibility. It must be a comfort for you to know that it is a trained philosopher whose finger is on the button.

What is wrong with this egomaniacal fantasy? Well, in a certain sense, nothing; barring whatever problems there may be about simultaneity, 'is H at *t*' and 'is T at *t*' are perfectly well defined predicates and they pick out perfectly well-defined (relational) properties of physical particles. Anybody who can get at my dime can, indeed, affect the distribution of these properties throughout the universe. It's a matter of temperament whether one finds it fun to do so.

What *would* be simply mad, however, would be to try to construct a particle physics which acknowledges *being an H-particle* or *being a T-particle* as part of its explanatory apparatus. *Why* would that be mad? Because particle physics, like every other branch of science, is in the business of causal explanation; and whether something is an H-(/T-) particle *is irrelevant to its causal powers*. I don't know exactly what that means; but whatever it means, I'm morally certain that it's true. I propose to wade around in it a bit.

Here are some things it seems to me safe to assume about science: We want science to give causal explanations of such things (events, whatever) in nature as can be causally explained.[2] Giving such explanations essentially involves projecting and confirming causal generalizations. And causal generalizations subsume the things they apply to in virtue of the causal properties of the things they apply to. Of course.

So what you need in order to do science is a taxonomic apparatus that distinguishes between things insofar as they have *different* causal properties, and that groups things together insofar as they have the *same* causal properties. So now we can say why it would be mad to embrace a taxonomy which takes seriously the difference between H-particles and T-particles. All else being equal, H-particles and T-particles have *identical* causal properties; whether something is an H-(T-) particle is

[2] There may be scientific enterprises that are not—or not primarily—interested in causal explanation; natural history, for example. And in these sciences, it is perhaps *not* identity and difference of causal powers that provides the criterion for taxonomic identity. But propositional attitude psychology is in the business of causal explanation or it is out of work; so this is a matter that we can afford to ignore.

irrelevant to its causal powers. To put it a little more tensely, if an event *e* is caused by H-particle *p*, then that same event *e* is also caused by *p* in the nearest nomologically possible world in which *p* is T rather than H. (If you prefer some other way of construing counterfactuals, you are welcome to substitute it here. I have no axes to grind.) So the properties of being H (/T) are taxonomically irrelevant for purposes of scientific causal explanation. But similarly for the properties of being H and T *brainstates*. And similarly for the properties of being H and T *mental states. And similarly for the property of being a mental state of a person who lives in a world where there is XYZ rather than H_2O in the puddles.* These sorts of differences in the relational properties of psychological (/brain/particle) states are *irrelevant to their causal powers*; hence irrelevant to scientific taxonomy.

So, to summarize, if you're interested in causal explanation, it would be mad to distinguish between Oscar's brain states and Oscar2's; their brain states have identical causal powers. That's why we individuate brain states nonrelationally. And, similarly, if you are interested in causal explanation, it would be mad to distinguish between Oscar's *mental* states and Oscar2's; their mental states have identical causal powers. But commonsense deploys a taxonomy which *does* distinguish between the mental states of Oscar and Osar2. So the commonsense taxonomy won't do for the purposes of psychology. QED.[3]

However, I can imagine somebody not being convinced by this argument. For the argument depends on assuming that the mental states of Twins do in fact have the same causal powers,

[3] The implication is that commonsense attitude attributions aren't—or, rather, aren't *solely*—in aid of causal explanation; and this appears to be true. One reason why you might want to know what Psmith believes is in order to predict how he will behave. But another reason is because beliefs are often *true*, so if you know what Psmith believes, you have some basis for inferring how the world is. The relevant property of Psmith's beliefs for this latter purpose, however, is not their causal powers but something like *what information they transmit* (see Dretske (1981)). And, quite generally, what information a thing transmits depends on relational properties of the thing which may not affect its causal powers. My utterance 'water is wet' has, let's say, the same causal powers as my Twin's; but—assuming that both utterances are true—one transmits the information that H_2O is wet and the other transmits the information that XYZ is.

It is, I think, the fact that attitude ascriptions serve both masters that is at the bottom of many of their logical peculiarities; of the pervasiveness of opacity/transparency ambiguities, for example.

and I can imagine somebody denying that this is so. Along either of the two following lines:

First line: 'Consider the effects of my utterances of the form of words "Bring water!". Such utterances normally eventuate in somebody bringing me *water*; viz. in somebody bringing me H_2O. Whereas, by contrast, when my Twin utters "Bring water!" what he normally gets is water2; viz. XYZ. So the causal powers of my water-utterances do, after all, differ from the causal powers of my Twin's "water"-utterances. And similarly, *mutatis mutandis*, for the causal powers of the mental states that such utterances express. And similarly, *mutatis mutandis*, for the mental states of the Oscars in respect of brisket and brisket2.'

Reply: This will *not* do; *identity of causal powers has to be assessed ACROSS contexts, not WITHIN contexts.*

Consider, if you will, the causal powers of your biceps and of mine. Roughly, our biceps have the *same* causal powers if the following is true: *for any thing x and any context C, if you can lift x in C, then so can I; and if I can lift x in C, then so can you.* What is, however, *not* in general relevant to comparisons between the causal powers of our biceps is this: that there is a thing x and a pair of contexts C and C′ such that you can lift x in C *and I can not lift x in C′*. Thus suppose, for example, that in C (a context in which this chair is *not* nailed to the floor) you can lift it; and in C′ (a context in which this chair *is* nailed to the floor) I cannot. That eventuality would give your biceps nothing to crow about. Your biceps—to repeat the moral—have cause for celebration only if they can lift xs *in contexts in which my biceps can't.*

Well, to return to the causal powers of the water utterances (/water thoughts) of Twins: It's true that when I say 'water' I get water and when my Twin says 'water' he gets XYZ. But that's irrelevant to the the question about identity of causal powers *because these utterances (/thoughts) are being imagined to occur in different contexts.* (Mine occur in a context in which the local potable is H_2O, his occur in a context in which the local potable is XYZ.) What *is* relevant to the question of identity of causal powers are the following counterfactuals: (a) If this utterance (/thought) had occurred in my context, it *would have had* the effects that my utterance (/thought) did have; and (b) if my utterance (/thought) had occurred in his context, it *would have*

had the effects that his utterance (/thought) did have. For our utterances (/thoughts) to have the same causal powers, both of those counterfactuals have to be true. But both of those counterfactuals *are* true since (for example) if I had said 'Bring water!' on Twin Earth, it's XYZ that my interlocutors would have brought; and if he had said 'Bring water!' here, his interlocutors would have brought him H_2O. So, OK so far; we have, so far, no reason to suppose that the causal powers of my Twin's mental states are different from the causal powers of mine.

Second line: 'Consider the *behavioral* consequences of the mental states of Oscar and Oscar2. (I assume here and throughout, that the interesting relations between behaviors and states of mind are typically causal. Philosophers have denied this, but they were wrong to do so.) Oscar's thoughts and desires sometimes eventuate in his *saying* such things as that he prefers brisket to, as it might be, hamburger; Oscar's thoughts sometimes lead to his evincing brisket eating preferences and brisket purchasing behavior; and so forth. Whereas, Oscar2 *never* does *any* of these things. Oscar2 may, of course, say that he likes brisket2; and he may evince brisket2 preferences; and he may, when appropriately stimulated by a meat counter, behave brisket2-purchasingly.[4] And, of course, when he says and does these things, he may produce precisely the same bodily *motions* as his counterpart produces when he says and does the corresponding things *vis à vis* brisket. But all that shows is that behaving isn't to be identified with moving one's body; a lesson we ought to have learned long ago.'

There's another aspect of this line of reply that's worth noticing: Independent of the present metaphysical issues, anybody who takes the Burge/Putnam intuitions to be decisive for the individuation of the attitudes has a strong motive for denying that Oscar and Oscar2's behavior (or Mine and My Twin's) are, in general, type-identical. After all, behavior is supposed to be the result of mental causes, *and you would generally expect different mental causes to eventuate in correspondingly different behavioral effects*. By assumption the Twins' attitudes (and the two Oscars') differ a lot, so if these very different sorts of mental

[4] Since all brisket2 is brisket (though not *vice versa*) every brisket2 purchase is a brisket purchase. This, however, is a consideration not profoundly relevant to the point at issue.

causes nevertheless invariably converge on identical behavioral effects, that would seem to be an accident on a very big scale. The way out is obviously to deny that the behavioral identity holds, to insist that the commonsense way of identifying behaviors, like the commonsense way of identifying the attitudes, goes out into the world for its principles of individuation; that it depends essentially on the relational properties of the behavior.

In short, Barbara Pym's question: 'Where does "behavior" begin and end?' is one that needs to be taken seriously in a discussion of the causal powers of mental states. Assuming, as indeed I have been doing, that My mental states and My Twin's are identical in causal powers begs that question; or so, in any event, the objection might go.

Reply: To begin with, you can, of course, make the same move in respect to H-particles and T-particles. Here's how it would sound: 'Being H rather than T does affect causal powers after all; for H-particles enter into H-particle interactions, and no T-particle does. H-particle interactions may, of course, *look* a lot like T-particle interactions, just as Oscar2's brisket2 eating behaviors look a lot like Oscar's brisket eating behaviors, and just as My water-requests sound a lot like my Twin's requests for XYZ. Philosophers are not, however, mislead by mere appearances; we see where the eye does not.

The least that all this shows is how taxonomic and ontological decisions intertwine: You can *save* classification by causal powers *come what may* by fiddling the criteria for event identity. To classify by causal powers is to count no property as taxonomically relevant unless it affects causal powers. But *x*'s having property P affects *x*'s causal powers just in case *x wouldn't have caused the same events* had it not been P. But, of course, whether *x would* have caused the same events had it not been P depends a lot on which events you count as the same and which you count as different. In the present case, whether the difference between being H and being T affects a particle's causal powers depends on whether the very same event which *was* an interaction of H-particles *could have been* an interaction of T-particles. (Perhaps it goes without saying that the principle that events are individuated by their causes and effects is perfectly useless here; we can't apply it unless we already know whether an event that *was*

caused by an H-particle could have had *the same cause* even if it had been the effect of a T-particle.)

Could it be that this is a dead end? It looked like the notion of taxonomy by causal powers gave us a sort of *a priori* argument for individualism, and thus put some teeth into the idea that a conception of mental state suitable for the psychologist's purposes would have to be interestingly different from the commonsense conception of a propositional attitude. But now it appears that the requirement that states with identical causal powers ought *ipso facto* to be taxonomically identical can be met *trivially* by anyone prepared to make the appropriate ontological adjustments. Yet surely there *has* to be something wrong here; because it's false that two events could differ *just* in that one involves H-particles and the other involves T-particles; and it's false that H-particles and T-particles differ in their causal powers; and—as previously noted—it would be *mad* to suggest saving the supervenience of the propositional attitudes by individuating brainstates relationally. And, moreover, it is very plausible that all these intutitions hang together. The question is: what on earth do they hang *on*?

I hope I have managed to make this all seem very puzzling; otherwise you won't be impressed when I tell you the answer. But in fact the mystery is hardly bigger than a bread box, and certainly no deeper. Let's go back to the clear case and trace it through.

If H-particle interactions are *ipso facto* different events from T-particle interactions, then H-particles and T-particles have different causal powers. But if H-particles and T-particles have different causal powers, then the causal powers—not just certain of the relational properties, mind you, but *the causal powers*—of every physical particle in the universe depend on the orientation of my gen-u-ine United States ten cent piece. That includes, of course, physical particles that are *a long way* away; physical particles on Alpha Centuri, for example. And *that's* what's crazy because, while such relational properties as being H or being T can depend on the orientation of my dime *by stipulation*, how on Earth could the *causal powers* of particles on Alpha Centuri depend on the orientation of my dime? Either there would have to be a causal mechanism to mediate this dependency, or it would have to be mediated by a fundamental

law of nature; and there aren't any such mechanisms and there aren't any such laws. *Of course* there aren't.

So, then, to avoid postulating impossible causal mechanisms and/or impossible natural laws, we will have to say that, all else being equal, H-particle interactions are *not* distinct events from T-particle interactions; hence that H-particles and T-particles do *not* differ in their causal powers; hence that the difference between being an H-particle and being a T-particle does *not* count as taxonomic for purposes of causal explanation. Which is, of course, just what intuition tells you that you *ought* to say.

Exactly the same considerations apply, however, to the individuation of mental states. If every instance of brisket-chewing behavior *ipso facto* counts as an event distinct from any instance of brisket2-chewing behavior, then, since brisket-cravings cause brisket-chewings and brisket2-cravings don't, Oscar's mental state differs in its causal powers from Oscar2's. But then there must be some mechanism which connects the causal powers of Oscar's mental states with the character of the speech community that he lives in *and which does so without affecting Oscar's physiology* (remember, Oscar and Oscar2 are molecularly identical). But there is no such mechanism; *you can't* affect the causal powers of a person's mental states without affecting his physiology (That's not a conceptual claim or a metaphysical claim. It's a contingent fact about how God made the world.) So, in order to avoid postulating crazy causal mechanisms, we have to assume that brisket chewings are not *ipso facto* events distinct from chewings of brisket2; hence that brisket cravings do not *ipso facto* have different causal powers from brisket2 cravings; hence that, for purposes of causal explanation, Oscar's cravings count as mental states of the same kind as Oscar2's.

There is, I think, no doubt but that we *do* count that way when we do psychology. Ned Block has a pretty example that makes this clear. He imagines a psychologist (call her 'Psyche'—the 'P' is silent, as in 'Psmith') who is studying the etiology of food preferences, and who happens to have both Oscar and Oscar2 in her subject population. Now, on the intuitions that Burge invites us to share, Oscar and Oscar2 have *different* food preferences; what Oscar prefers to gruel is brisket, but what Oscar2 prefers to gruel is brisket2. Psyche, being a proper psychologist, is of course

interested in sources of variance; so that the present case puts Psyche in a pickle. If she discounts Oscar and Oscar2, she'll be able to say—as it might be—that there are two determinants of food preference: 27.3% of the variance is genetic and the remaining 72.7% is the result of early training. If, however, she counts Oscar and Oscar2 in, and if she counts their food preferences the way that Burge wants her to, then she has to say that there are *three* sources of variance: genetic endowment, early training *and linguistic affiliation.* But surely, it's *mad* to say that linguistic affiliation is *per se* a determinant of food preference; how *could* it be?[5]

I think it's perfectly clear how Psyche out to jump: she ought to say that Oscar and Oscar2 count as having *the same* food preferences and therefore do *not* constitute counterexamples to her claim that the determinants of food preference are exhausted by genes and early training. And the previous discussion makes clear just *why* she ought to say this: if Oscar and Oscar2 have different food preferences, then there must be some difference in the causal powers of their mental states—psychological taxonomy is taxonomy *by* causal powers. But if there is such a difference, then there must be some mechanism which can connect the causal powers of Oscar's mental state with the character of his linguistic affiliation *without affecting his physiological constitution.* But there is no such mechanism; the causal powers of Oscar's mental states supervene on his physiology, just like the causal powers of your mental states and mine.

Well, if all this is as patent as I'm making it out to be, how could anyone have ever supposed that the standards of individuation appropriate to the psychologist's purposes are

[5] Burge points out (personal communication) that the Oscars' food preferences *don't* differ if you individuate *de re*; i.e. that brisket and gruel are such that *both Oscars* prefer dining on the former to dining on the latter (a fact that Psyche could establish by testing them on samples). But I don't see that this helps since it seems to me thoroughly implausible that linguistic affiliation *per se* determines food preferences *de dicto.*

If it does, that opens up new vistas in nonintrusive therapy. For example, it looks as though we can relieve Oscar's unnatural craving for brisket just by changing the linguistic background; viz. by getting his colinguals to talk English2 instead of English. Whereas, it used to seem that we'd be required to operate on *Oscar*: desensitization training, depth therapy, Lord knows what all else.

Psyche and I find this sort of consequence preposterous, but no doubt intuitions differ. That's why it's nice to have a principle or two to hone them on.

other than individualistic? I cast no aspersions, but I have a dark suspicion; I think people get confused as between methodological *individualism* and methodological *solipsism*. A brief excursus on this topic, therefore, will round off this part of the discussion.

Methodological individualism is the doctrine that psychological states are individuated *with respect to their causal powers*. Methodological solipsism is the doctrine that psychological states are individuated *without respect to their semantic evaluation*.[6]

Now, the semantic evaluation of a mental state depends on certain of its relational properties (in effect, on how the state corresponds to the world). So we could say, as a rough way of talking, that solipsistic individuation is *nonrelational*.

But if we are going to talk that way, then *it is very important* to distinguish between solipsism and individualism. In particular, though it's a point of definition that *solipsistic* individuation is *non*relational, there is nothing to stop principles of individuation from being simultaneously relational and individualistic. *Individualism does not prohibit the relational individuation of mental states*; it just says that no property of mental states, relational or otherwise, counts taxonomically unless it affects causal powers.

Indeed, individualism *couldn't* rule out relational individuation *per se* if any of what I've been arguing for up till now is true. I've taken it that individualism is a completely general methodogical principle in science; one which follows simply from the scientist's goal of causal explanation and which, therefore, all scientific taxonomies must obey. By contrast, it's patent that taxonomic categories in science are *often* relational. Just as you'd expect, relational properties can count taxonomically whenever they effect causal powers. Thus 'being a planet' is a relational property *par excellence*, but it's one that individualism permits to operate in astronomical taxonomy. For whether you are a planet affects your trajectory and your trajectory determines what you can bump into; so whether you're a planet affects your

[6] More precisely, methodological solipsism is a doctrine—not about individuation in psychology at large but—about individuation in aid of the psychology of mental processes. Methodological solipsism constrains the ways mental processes can specify their ranges and domains: They can't apply differently to mental states just in virtue of the truth or falsity of the propositions that the states express. And they can't apply differently to concepts depending on whether or not the concepts denote. (See Fodor 1978) This is, however, a nicety that is almost always ignored in the literature and I shan't bother about it here.

causal powers, which is all that individualism asks for. Equivalently: the property of being a planet is taxonomic because there are causal laws that things satisfy in virtue of being planets. By contrast, the property of living in a world in which there is XYZ in the puddles is *not* taxonomic because there are *no* causal laws that things satisfy in virtue of having *that* property. And similarly for the property of living in a speech community in which people use 'brisket' to refer to brisket of beef. The operative consideration is, of course, that where there are no causal laws about a property, having the property has no effect on causal powers.

To put the point the other way around, solipsism (construed as prohibiting the relational taxonomy of mental states) is unlike individualism in that it *couldn't conceivably* follow from any *general* considerations about scientific goals or practices. 'Methodological solipsism' is, in fact, an empirical theory about the mind: it's the theory that mental processes are computational, hence syntactic. I think this theory is defensible; in fact, I think it's true. But its defence can't be conducted on *a priori* or metaphysical grounds and its truth depends simply on the facts about how the mind works. Methodological solipsism differs from methodological individualism in both these respects.

Well, to come to the point: if you happen to have confused individualism with solipsism (and if you take solipsism as the doctrine that psychological taxonomy is nonrelational) then you might try arguing against individualism by remarking that the psychologist's taxonomic apparatus is, often enough, nonsolipsistic (viz. that it's often relational). As, indeed, it is. Even computational ('information flow') psychologists are professionally interested in such questions as: 'Why does this organism have the computational capacities that it has?; Why does its brain compute this algorithm rather than some other?' or even 'Why is this mental process generally truth preserving?' Such questions often get answered by reference to relational properties of the organism's mental state. See for example Ullman (1979) where you get lovely arguments that run like this: 'This perceptual algorithm is generally truth preserving because the organism that computes it lives in a world where most spatial transformations of objects are rigid. If the same algorithm were run in a world in which most spatial transformations were not

rigid, it wouldn't be truth preserving, and the ability to compute it would be without survival value. So, presumably, the organism wouldn't have this ability in such a world.' These sorts of explanations square with *individualism*, because the relational facts they advert to affect the causal powers of mental states; indeed, they affect their very existence. But, naturally, explanations of this sort—for that matter, *all* teleological explanations—are *ipso facto* nonsolipsistic. So *if* you have confused solipsistic (viz. nonrelational) taxonomies with individualistic taxonomies (viz. taxonomies by causal powers) then you *might* wrongly suppose that the affection psychologists have for teleological explanation argues that they—like the laity—are prone to individuate mental states nonindividualistically. But it doesn't. And they aren't.

Well, I've gotten us where I promised to; back to where we started. There is a difference between the way psychology individuates mental states and the way that commonsense does. At least there is if you assume that the Burge/Putnam intuitions are reliable.[7] But this fact isn't, in and of itself, really very interesting; scientific taxonomy is forever cross-cutting categories of everyday employment. For that matter, the sciences are forever cross-cutting one another's taxonomies. Chemistry doesn't care about the distinction between streams and oceans; but geology does. Physics doesn't care about the distinction between bankers and butchers; but sociology does. (For that matter, physics doesn't care about the distinction between The Sun and Alpha Centuri either; sublime indifference!) None of this is surprising; things in Nature overlap in their causal powers to various degrees and in various respects; the sciences play these overlaps, each in its own way.

And, for nonscientific purposes, we are often interested in taxonomies that cross cut causal powers. Causal explanation is just one human preoccupation among many; individualism is a

[7] It is, however, worth echoing an important point that Burge makes; the differences between the way that these taxonomies carve things up only show in funny cases. In practically all the cases that anybody actually encounters outside philosophical fantasies, the states that one is tempted to count as token beliefs that P share not just the causal powers that psychologists care about, but also the relational background to which the commonsense taxonomy is sensitive. This enormous *de facto* coextension is part of the argument that the psychologist's story really is a vindication of the commonsense belief/desire theory.

constitutive principle of *science*, not of rational taxonomy *per se*. Or, to put it a little differently—more in the material mode—God could make a genuine electron, or diamond, or tiger, or person because being an electron or a diamond or a tiger or a person isn't a matter of being the effect of the right kind of causes; rather, it's a matter of being the cause of the right kind of effects. And similarly, I think, for all the other natural kinds. Causal powers are decisively relevant to a taxonomy of natural kinds because such taxonomies are organized in behalf of causal explanation. Not all taxonomies have that end in view, however, so not all taxonomies classify by causal powers. Even God couldn't make a gen-u-ine United States ten cent piece; *only* the U.S. Treasury Department can do that.

You can't, in short, make skepticism just out of the fact that the commonsense way of taxonomizing the mental differs from the psychologist's way. You might, however, try the idea that disagreement between the commonsense taxonomy and the scientific one matters more in psychology than it does elsewhere *because psychology needs the commonsense notion of mental content*. In particular, you might try the idea that the notion of mental content doesn't survive the transition from the layman's categories to the scientist's. I know of at least one argument that runs that way. Let's have a look at it.

What we have—though only by assumption, to be sure—is a typology for mental states according to which My thoughts and my Twin's (and Oscar's thoughts and Oscar2's) have identical contents. More generally, we have assumed a typology according to which the physiological identity of organisms guarantees the identity of their mental states (and, *a fortiori*, the identity of the contents of their mental states). All this is entailed by the principle—now taken to be operative—that the mental supervenes upon the physiological together with the assumption—which I suppose to be untendentious—that mental states have their contents essentially, so that typological identity of the former guarantees typological identity of the latter. Alright so far.

But now it appears that even if the physiological identity of organisms ensures the identity of their mental states and the identity of mental states ensures the identity of contents, *the identity of the contents of mental states does not ensure the identity of their extensions:* My thoughts and my Twin's—like Oscar and

Oscar2's—*differ in their truth conditions* so it's an accident if they happen to have the same truth values. Whereas what makes my water-thoughts true is the facts about H_2O, what makes my Twin's 'water'-thoughts true is the facts about XYZ. Whereas the thought that I have—when it runs through my head that water is wet—is true iff H_2O is wet, the thought that he has—when it runs through his head that 'water' is wet—is true iff XYZ is wet. And it's an accident (it's just contingent) that H_2O is wet iff XYZ is. (Similarly, what I'm thinking about when I think: *water*, is different from what he's thinking about when he thinks: '*water*'; he's thinking about XYZ but I'm thinking about H_2O. So the denotations of our thoughts differ.) Hence, the classical—Putnamian—formulation of the puzzle about Twins: if mental state supervenes upon physiology, then thoughts don't have their truth conditions essentially; two tokens of the *same* thought can have *different* truth conditions, hence different truth values. If thoughts are in the head, then content doesn't determine extension.

That, then, is the 'Twin-Earth Problem'. Except that so far it *isn't* a problem; it's just a handful of intuitions together with a commentary on some immediate implications of accepting them. If that were *all*, the right response would surely be 'So what?'. What connects the intuitions and their implications with the proposal that we give up on propositional attitude psychology is a certain *Diagnosis*. And, while a lot has been written about the intuitions and their implications, the diagnosis has gone largely unexamined. I propose now to examine it.

Here's the Diagnosis

'Look, on *anybody's* story, the notion of content has got to be at least a little problematic. For one thing, it seems to be a notion *proprietary* to the information sciences, and *soi-disant* 'emergents' bear the burden of proof. At a minimum, if you're going to have mental contents, you owe us some sort of account of their individuation.

'Now, prior to the Twin-Earth problem, there *was* some sort of account of their individuation; you could say, to a first approximation, that identity of content depends on identity of extension. No doubt that story leaked a bit: Morning-Star thoughts look to be different in content from the corresponding

Evening-Star thoughts, even though their truth conditions are arguably the same. But at least one could hold firmly to this: 'Extension supervenes on content; no difference in extension without some difference in content.' Conversely, it was a *test* for identity of content that the extensions had to come out to be the same. And that was the *best* test we had; it was the one source of evidence about content identity that seemed *surely reliable.* Compare the notorious wobbliness of intuitions about synonymy, analyticity and the like.

'But now we see that *it's not true after all* that differences of extension implies difference of content; so unclear are we now about what content-identity come to—hence about what identity of propositional attitudes comes to—that we can't even assume that typologically identical thoughts will always be true and false together. The consequence of the psychologist's insistence on preserving supervenience is that *we now have no idea at all* what criteria of individuation for propositional attitudes might be like; hence we have *no idea at all* what counts as *evidence* for the identity of propositional attitudes.

'Short form: Inferences from difference of extension to difference of content used to bear almost all the weight of propositional attitude attribution. That was, however, a frail reed and now it has broken. The Twin-Earth Problem *is* a problem *because it breaks the connection between extensional identity and content identity.*'

Now, the Twin-Earth intuitions are fascinating, and if you care about semantics you will, no doubt, do well to attend to them. But, as I've taken pains to emphasize, you need the Diagnosis to connect the intuitions about Twins to the issues about the facticity of belief/desire psychology, and—fortunately for those of us who envision a psychology of propositional attitudes—the Diagnosis rests on a quite trivial mistake: *the twin-earth examples don't* break *the connection between content and extension; they just relativize it to context.*

Suppose that what you used to think, prior to Twin-Earth, is that contents are something like functions from thoughts to truth conditions: given the content of a thought, you know the conditions under which that thought would be true. (Presumably a truth condition would itself then be a function from worlds to truth values: a thought that has the truth condition TC takes the

value T in world W iff TC is satisifed in W. So, for example: in virtue of its content, the thought that it's raining has the truth condition *that it's raining* and is thus true in a world iff it's raining in that world.) I hasten to emphasize that if you don't—or didn't—like that story, it's quite alright for you to choose some other; my point is going to be that if you liked *any story of that kind* before Twin-Earth, you're perfectly free to go on liking it now. For, even if all the intuitions about Twin-Earth are right, and even if they have the implications that they are said to have, extensional identity still constrains intentional identity because *contents still determine extensions relative to a context*. If you like, contents are functions from *contexts* and thoughts onto truth conditions.

What, if anything, does that mean? Well, there is presumably something about the relation between Twin-Earth and Twin-Me in virtue of which his 'water'-thoughts are about XYZ even though my water-thoughts are not. Call this condition that's satisfied by <Twin-Me, Twin-Earth> condition C (because it determines the Context of his 'water'-thoughts). Similarly, there must be something about the relation between me and Earth in vritue of which my water-thoughts are about H_2O even though my Twin's 'water'-thoughts are not. Call this condition that is satisfied by <me, Earth> condition C′. I don't want to worry, just now, about the problem of how to articulate conditions C and C′. Some story about constraints on the causal relations between H_2O tokenings and water-thought tokenings (and between XYZ tokenings and 'water'-thought tokenings) would be the obvious proposal; but it doesn't matter much for the purposes now at hand. Because we *do* know this: Short of a miracle, it must be true that if an organism shares the neurophysical constitution of my Twin *and satisfies C*, it follows that its thoughts and my Twin's thoughts share their truth conditions. For example, short of a miracle the following counterfactual must be true: given the neurological identity between us, in a world where I am in my Twin's context my 'water'-thoughts are about XYZ iff his are. (And, of course, vice versa: in a world in which my Twin is in my context, given the neurological identity between us, it must be that his water thoughts are about H_2O iff mine are.)

But now we have an extensional identity criterion for mental contents: two thought contents are identical only if they effect the same mapping of thoughts and contexts onto truth

conditions. Specifically, your thought is content-identical to mine only if in every context in which your thought has truth condition T, mine has truth condition T and vice versa.

It's worth re-emphasizing that, by this criterion, my Twin's 'water'-thoughts are intentionally identical to my water-thoughts; they have the same contents even though, since their contexts are *de facto* different, they differ, *de facto*, in their truth conditions. In effect, what we have here is an extensional criterion for 'narrow' content (see above). The 'broad content' of a thought, by contrast, is what you can semantically evaluate; it's what you get when you specify a narrow content *and fix a context*. This makes the notion of narrow content the more basic of the two; which is just what sensible people have always supposed it to be.

We can now see why we ought to reject both of the following two suggestions found in Putnam (1975): That we consider the extension of a term (/concept/thought) to be an independent component of its 'meaning vector'; and that we make do, in our psychology, with stereotypes *instead of* contents. The first proposal is redundant since, as we've just seen, contents (meanings) determine extensions given a context. The second proposal is unacceptable because, unlike contents, stereotypes *don't* determine extensions *even* given a context. (Since it's untendentious that stereotypes supervene on physiology, the stereotypes for real water and Twin-water must be identical.) But, as the Diagnosis rightly says, we need an extension-determiner as a component of the meaning vector because we rely on 'different extension different content' for the individuation of concepts.

—Stop, stop! I have an objection.

—Sing me your song, Oh!

—Well, since, on your view, your water-thoughts are content identical to your Twin's, I suppose we may infer that the English word 'water' has the same intension as its Tw-English homonym (hereinafter spelled 'water2')?

—We may.

—'But if 'water' and 'water2' have the same intentions, they must apply to the same things. So since 'water2' applies to XYZ, 'water' applies to XYZ too. It follows that XYZ must *be* water (what else could it mean to say that 'water' applies to it?) But, as

a matter of fact, XYZ *isn't* water; only H_2O is water. Scientists discover essences.

—I don't know whether scientists discover essences. It may be that philosophers make them up. In either event, the present problem doesn't exist. The denotation of 'water' is determined not just by its meaning but by its context. But the context for English 'anchors' 'water' to H_2O just as, *mutatis mutandis*, the context for Tw-English anchors 'water2' to XYZ. (I learned 'anchors' at Stanford; it is a very useful term despite—or maybe because of—not being very well-defined. For present purposes, an expression is anchored iff it has a determinate semantic value.) So then, the condition for *x* is 'water' to be true requires that *x* be H_2O. Which, by assumption, XYZ isn't. So English 'water' doesn't apply to XYZ (though, of course, Tw-English 'water' does). OK so far.

And yet . . . and yet! One seems to hear a Still Small Voice—could it be the voice of conscience?—crying out as follows: You say that 'water' and its Tw-English homonym mean the same thing; well then *what* do they mean?

How like the voice of conscience to insist upon the formal mode. It might equally have put its problem this way: 'What *is* the thought such that when I have it its truth condition is that H_2O is wet and when my Twin has it its truth condition is that XYZ is wet? What is the concept *water* such that it denotes H_2O in this world and XYZ in the next?' I suspect that this—and not Putnam's puzzle about individuation—is what *really* bugs people about narrow content. The construct invites a question which—so it appears—we simply don't have a way of answering.

But, conscience be hanged, the question is radically ill advised. What the Still Small Voice wants me to do is utter an English sentence which expresses just what my 'water'-thoughts have in common with my Twin's. Unsurprisingly, I can't do it. That's because the content that an English sentence expresses is *ipso facto anchored* content, hence *ipso facto not* narrow.

So, in particular, qua expression of English 'water is wet' is anchored to the wetness of water (i.e. of H_2O) just as, qua expression of Tw-English, 'water2 is wet' is anchored to the wetness of water2 (i.e. to the wetness of XYZ). And, of course, since it is anchored to water, 'water is wet' doesn't—can't—express the narrow content that my water-thoughts share with

my Twin's. Indeed, if you mean by content what can be semantically evaluated, then what my water-thoughts share with Twin 'water'-thoughts *isn't* content. Narrow content is radically inexpressible because it's only content *potentially*; it's what gets to *be* content when—and only when—it gets to be anchored. We can't—to put it in a nutshell—say what Twin thoughts have in common. This is because what can be said is *ipso facto* semantically evaluable, and what Twin-thoughts have in common is *ipso facto* not.

Here is another way to put what is much the same point: You have to be sort of careful if you propose to co-opt the notion of narrow content for service in a 'Gricean' theory of meaning. According to Gricean theories, the meaning of a sentence is inherited from the content of the propositional attitude(s) that the sentence is conventionally used to express. Well, that's fine so long as you remember that it's *anchored* content (that is, it's the content of anchored attitudes), and hence not narrow content, that sentences inherit. Looked at the other way around, when we use the content of a sentence to specify the content of a mental state (viz. by embedding the sentence to a verb of propositional attitude) the best we can do—in principle, *all* we can do—is avail ourselves of the content of the sentence qua anchored; for it's only qua anchored that sentences *have* content. The corresponding consideration is relatively transparent in the case of demonstratives. Suppose the thought 'I've a sore toe' runs through your head and also runs through mine; what's the content that these thoughts share? Answer: *you can't say what it is by using a sentence* since, whenever you use a sentence that contains 'I', the 'I' that it contains automatically gets anchored to you. You can, however, sneak up on the shared content by *mentioning* a sentence, as I did just above. In such cases, mentioning a sentence is a way of abstracting a form of words from the consequences of its being anchored.

One wants, above all, to avoid a sort of fallacy of subtraction: 'Start with anchored content; take the anchoring conditions away, and you end up with a *new sort of content*, an unanchored content; a *narrow* content, as we say.' (Compare: 'start with a bachelor; take the unmarriedness away, and you end up with *a new sort of bachelor*, a married bachelor; a *narrow* bachelor, as we say.') Or, rather, there's nothing wrong with talking that way, so

long as you don't then start to wonder *what the narrow content of—for example—the thought that water is wet could be*. Such questions can't be answered in the nature of things; so, in the nature of things, they shouldn't be asked.[8] People who positively *insist* on asking them generally get what they deserve: phenomonalism, verificationism, 'procedural' semantics or skepticism, depending on temperament and circumstance.

—'But look' the SSV replies, 'if narrow content isn't really content, then in what sense do you and your Twin have any water thoughts in common at all? And if the form of words "water is wet" doesn't express the narrow content of Twin water-thoughts, how can the form of words "the thought that water is wet" succeed in picking out a content that your thoughts share with your Twin's?'

—Answer: What I share with my Twin—what supervenience *guarantees* that we share—is a mental state that is semantically evaluable relative to a context. Referring expressions of English can therefore be used to pick out narrow contents via their *hypothetical* semantic properties. So, for example, the English expression: 'the thought that water is wet' can be used to specify the narrow content of a mental state that my Twin and I share (even though, qua anchored to H_2O, it doesn't, of course, *express* that content). In particular, it can be used to pick out the content of my Twin's 'water' thought via the truth conditions that it *would have had* if my Twin had been plugged into my world. Roughly speaking, this tactic works because the narrow thought that water is wet is the *unique* narrow thought that yields

[8] Since you buy the narrow content construct at the cost of acknowledging a certain amount of inexpressibility, it may be some consolation that *not* buying the narrow content construct also has a certain cost in inexpressibility (though for quite a different sort of reason, to be sure). So, suppose you think that Twin-Earth shows that content doesn't determine extension and/or that content doesn't supervene on physiology. So, you have no use for narrow content. Still there's the following question: When my Twin thinks 'water2 is wet', how do you say, in English, what he is thinking? Not, by saying 'water2 is wet' for that's a sentence of Tw-English; and not by saying 'water is wet' since, on the present assumption, whatever 'water2' means, it's something different from what 'water' means; not by saying 'XYZ is wet', since my Twin will presumably take 'water2 is XYZ' to say something informative; something, indeed, which he might wish to deny. And not, for sure, by saying 'H_2O is wet' since there isn't any H_2O on Twin Earth, and my Twin has never so much as heard of the stuff. It looks like the meaning of 'water2 is wet' is *inexpressible* in English. And, of course, the same thing goes—only the other way round—for expressing the meaning of 'water' in Tw-English.

the truth condition *H_2O is wet* when anchored to my context and the truth condition *XYZ is wet* when anchored to his.

You can't, in absolute strictness, *express* narrow content; but as we've seen, there are ways of sneaking up on it.

—SSV: 'By that logic, why don't you call the narrow thought you share with your Twin "the thought that water2 is wet"? After all, that's the 'water-thought' that you would have had if you had been plugged into your Twin's context (and that he *does have* in virtue of the fact that *he has* been plugged into his context.) Turn about is fair play.'

Answer: (a) 'the thought that water2 is wet' is an expression of Tw-English; I don't speak Tw-English. (b) The home team gets to name the intension; the actual word has privileges that merely counterfactual worlds don't share.

—SSV: What about if you are a brain in a vat? What about then?

Answer: If you are a brain in a vat, then you have, no doubt, got serious cause for complaint. But it may be some consolation that brains in vats have no special *semantical* difficulties according to the present account. They are, in fact, just special cases of Twins.

On the one hand, a brain in a vat instantiates the same function from Contexts to truth conditions that the corresponding brain in a head does; being in a vat does not, therefore, affect the narrow content of one's thoughts. On the other hand, it *may* affect the *broad* content of one's thoughts; it may, for example, affect their truth conditions. That would depend on just which kind of brain-in-a-vat you have in mind; for example, on just what sort of connections you imagine that there are between the brain, the vat, and the world. If you imagine a brain in a vat that's hooked up to *this* world, and hooked up *just* the same way that one's own brain is, then—of course—that brain shares one's thought-contents *both* narrow *and* broad. Broad content supervenes on neural state together with connections to context. It had better, after all; a skull is kind of vat too.

—SSV: I do believe that you've gone over to Steve Stich. Have you no conscience? Do you take me for a mere expository convention?

—There, there; don't fret! What is emerging here is, in a certain sense, a 'no content' account of narrow content; but it is

nevertheless also a fully intentionalist account. According to the present story, a narrow content is *essentially* a function from contexts onto truth conditions; different functions from contexts onto truth conditions are *ipso facto* different narrow contents. It's hard to see what more you could want of an intensional state than that it should have semantic properties that are intrinsic to its individuation. In effect, I'm prepared to give Stich everything except what he wants.

Now, sleep conscience!

* * *

What I hope this discussion has shown is this: Given the causal explanation of behavior as the psychologist's end in view, he has motivation for adopting a taxonomy of mental states that respects supervenience. However, the psychologist needs a way to reconcile his respect for supervenience with the idea that the extension of a mental state constrains its content; for he needs to hold onto the argument from *difference* of extension to *difference* of content. When it comes to individuating mental states, that's the best kind of argument he's got, just as Putnam says. It turns out, however, that it's not hard to reconcile respecting supervenience with observing extensional constraints on content because you can relativize the constraints to context: given a context, contents are different if extensions are. There isn't a shred of evidence to suggest that this principle is untrue—surely the Twin cases propose no such evidence—or that it constrains content attributions any less well than the old, unrelativized account used to do. So it looks as though everything is alright. Let, therefore, rejoicing be unconstrained!

REFERENCES

Dretske, F. I. (1981): *Knowledge and the Flow of Information*, MIT Press, Cambridge Mass.

Fodor, J. (1978): 'Methodological Solipsism', in *Representations*, MIT Press, Cambridge Mass. 1981.

Putnam, H. (1975): 'The Meaning of "Meaning"', in K. Gunderson, ed., *Minnesota Studies in the Philosophy of Science*, University of Minnesota Press, p. 131-193.

Ullman, S. (1979): *The Interpretation of Visual Motion*, MIT Press, Cambridge Mass.

INDIVIDUALISM AND SUPERVENIENCE

Jerry Fodor and Martin Davies

II—Martin Davies

EXTERNALITY, PSYCHOLOGICAL EXPLANATION, AND NARROW CONTENT

In a massively influential body of papers and books, Jerry Fodor has urged the view that a scientific psychology will be computational, representational, and recognisably a precisification of the commonsense scheme of propositional attitude attribution and explanation. In 'Individualism and Supervenience' he seeks—not for the first time—to defend his position against a difficulty that is supposed to arise out of Putnam's 'The Meaning of "Meaning" ' ([1975])—the Twin Earth examples—and Burge's 'Individualism and the Mental' ([1979])—the 'brisket', 'arthritis', and 'sofa' examples, among many others.

The argumentative architecture of Fodor's paper is as follows. First, for the purposes of the argument, Fodor grants that Burge's examples extend to all concepts what Putnam's examples demonstrated about natural kind concepts, namely that the contents of attitudes involving those concepts are individuated in part by features that are *external* both to the attitude states and to the subjects of those states. *Prima facie*, this means that the commonsense scheme individuates contents in a way that will not do for a scientific psychology, since it does not respect the supervenience of the psychological upon the neurophysiological. In short, there is a *prima facie* mismatch between the commonsense scheme and the requirements of science. Second, the suggestion that a scientific psychology will itself individuate the contents of states by external features is indefensible. A scientific notion of content must respect supervenience, and the *prima facie* mismatch is a genuine mismatch. Third, we then face the apparent problem that, for any notion of content that is available to scientific psychology, content will not determine truth conditions. Consequently, we cannot infer from difference of truth conditions to difference of content, and we seem to lose our grasp on the very notion of content. But, fourth, this problem is only apparent, since

content *will* determine truth conditions relative to a context, so that we can infer from difference of truth conditions in the same context to difference of content. This narrow content is strictly speaking inexpressible; but it is still genuinely intensional. So, in sum, the prospects for a scientific psychology based on propositional attitudes are no worse than before the Putnam and Burge examples.

My discussion of Fodor's argument is in two parts. In the first part, I focus on the second stage of the argument. It is there that Fodor argues that the *prima facie* mismatch between the commonsense scheme and any notion of content available to science is a genuine mismatch. In the second part, I take up one aspect of the fourth stage of the argument, where Fodor argues that the apparent problem for a content based scientific psychology is readily solved by the narrow content construct.

I

1. In my sketch of the argument, I used the rather vague term 'external'. This is not Fodor's own term, and my use of it blurs some of the details of Fodor's position. For it is quite crucial to a proper understanding of the argument that one should distinguish two different contrasts that Fodor draws: nonrelational *vs* relational (or solipsistic *vs* non-solipsistic), and individualistic *vs* non-individualistic. (Cf. Fodor [1980].)

Initially, a classification of psychological states which classifies together the states of neurophysiologically identical twins is said to be *individualistic* (p. 238). That leads one to expect that what is constitutive of an individualistic classification is that it respects local supervenience—that is, the requirement that the psychological states of an individual supervene upon the neurophysiological, and ultimately upon the physical, states of that individual. Officially, however, individualistic classification is classification by causal powers, and it is explicit that classification by causal powers can, in general, answer to relational properties provided that those properties affect causal powers. So it is not immediate that individualistic classification in psychology respects local supervenience.

Indeed, on Fodor's account it seems that there will be cases of classification within psychology that do not respect local

supervenience. A relational property is allowed to count in an individualistic classification of a state provided that it affects the causal powers of the state. If having relational property *R* is part of the *function* of a particular mechanism—if it has been *selected for* its having *R*—then property *R* can count in an individualistic taxonomy. For having *R* affects the causal powers of the mechanism; it affects *its very existence* (p. 252).

Now, where we find this teleologically based relational, but still individualistic, classification in psychology, does that classification respect the supervenience of the psychological upon the neurophysiological? The question is not easy. After distinguishing between individualistic and nonrelational taxonomy, Fodor says that he takes to be operative 'the principle . . . that the mental supervenes upon the physiological' (p. 253). But it is far from obvious that the use of a teleologically based relational taxonomy in psychology requires a correspondingly relational physiological taxonomy of brain states. And to the extent that neurophysiological classification is nonrelational, then the principle of local supervenience would seem not to be quite accurate.

Suppose we have a mechanism with the relational property that it has a range of possible states which covary with a range of states in the world. The mechanism registers information about the world, and that is what it is for; a mechanism of that (psychological) type is present because mechanisms of that type register that information. The idea is that the relational property of the mechanism can count taxonomically in a psychological theory. There are then two kinds of example that need to be considered. As examples of the failure of supervenience, one is more likely to be convincing than the other.

In the first kind of example, we imagine a molecule for molecule twin of our mechanism with the *same* relational property, but where that relational property is *not* teleologically significant. Suppose, for example, that the existence of the mechanism is entirely fortuitous—like those copies of the *Encyclopedia Britannica* that result from explosions in printing factories. Then the relational property should not count taxonomically, and so this mechanism, though a physiological twin of the original, should be classified differently at the psychological level of description, thus infringing supervenience. This first kind kind of example is not likely to be convincing,

since it is notoriously difficult to sustain a firm intuition that the twin mechanism does not have the same function as the original. (On this kind of difficulty, see my [1983].)

In the second kind of example, we imagine a twin mechanism with a *different* relational property, but one which *is* also teleologically significant. We suppose that the creature housing the mechanism is set in a quite different external environment. The function of the mechanism is once again to register information about the external environment, but information of a quite different kind, with a quite different role in the experience and behaviour of the creature. The two mechanisms are physiologically twins, but would surely not be classified together for the purposes of scientific psychology. In short, it seems that relational but individualistic taxonomy is liable to violate supervenience.

How widespread is relational taxonomy in psychology? On Fodor's account, a relational property can count taxonomically when it is teleologically significant. But there are two different ways of taking this element of relationalism. Taken strictly, it might mean that one can advert to a relational property when one is actually engaged in providing a teleological explanation. Taken generously, it might mean that one can make free use of a relational description of a state or mechanism provided that the description *would* figure in a teleological explanation. I shall opt for the strict construal of Fodor's account, since that minimises the deviation from the principle of supervenience. As we shall shortly see, the generous construal would obscure the difference between Fodor's position and the position he is opposing in the second stage of his argument.

2. With so much by way of clarification of the distinction between nonrelational (solipsistic) and individualistic taxonomy, let us return to the sketch of Fodor's argument.

The first stage of the argument largely consists in accepting the claims of Putnam and Burge that the commonsense scheme individuates mental states in part by external factors, or—what comes to the same thing, given that mental states have their contents essentially—that the commonsense scheme makes use of an externalist notion of content. These are doctrines of what some would be pleased to call the *externality of the mind.* In order

for this to present the threat of a mismatch between the commonsense scheme and a scientific psychology, it must be the case that psychology does not, or at least should not, proceed in a similarly externalist spirit. This claim about the nature of psychology is what the second stage of the argument sets out to establish.

One of the targets of this stage is Burge who, in 'Individualism and Psychology' ([1986a]), defends at length the view that contemporary psychology is shot through with externalism. This view Burge would express by saying that psychology is not *individualistic*: for this term in Burge's hands is roughly equivalent to Fodor's 'nonrelational'. 'Individual' contrasts with both 'social' and 'environmental', and Burge's claim that contemporary cognitive psychology frequently describes states in ways that advert to the environment in which the subject is set (or is normally set) is quite compatible with a familiar claim that is *not* presently under discussion, namely that contemporary cognitive psychology is not social. The example of a psychological theory on which Burge focuses is Marr's theory of vision. (See Marr [1982].)

At the top level of Marr's hierarchy of levels of explanation, the computational task is described and is subject to mathematical and evolutionary investigation: From what other information could information about shape or depth be computed, and how could it be computed?; and Why would it be adaptive for the organism to be able to extract that information? At a lower level in the hierarchy, algorithms are postulated for performing subtasks of the overall computational task. These algorithms must be both neurophysiologically and evolutionarily plausible.

On the face of it, this theory is rampantly externalist. The description of the computational task is in externalist terms, with free use of notions such as objective (or object-centred) shape, and location (or depth). And since the algorithms are postulated as algorithms for performing elements of that task, the externalism flows down to the lower level in the hierarchy. What is more, the postulated algorithms are often tractably simple only because they are not fully general; and the simplifications are illuminatingly described in the context of the total theory by saying that the algorithms build in substantive assumptions about the world (about 'smoothness' of surfaces,

and 'rigidity' of transformations of objects), and would not be reliable in a world that did not measure up to those assumptions.

None of this externalism would be obviously incompatible with Fodor's position if we opted for the generous construal of the element of relationalism discussed in the last section. The externalist characterisation of the assumptions implicit in an algorithm is part of an externalist description of what the algorithm is computing. And all of that would be compatible with Fodor's individualism (though not, of course, with nonrelationalism—Burge's individualism), given the generous construal. For, as he points out, the implicit assumptions, for example, are teleologically significant.

Just how far Fodor's externalism would then extend is not perhaps obvious. Externalist taxonomy, and externalist—truth conditional—content in psychological theory, would need to be teleologically based in order to be legitimate by the lights of Fodor's individualism. So, roughly speaking, externalism in psychology would extend as far as there was a teleological component in the theory of truth conditional content. For a friend of causal-cum-teleological theories of truth conditions, like Fodor, that would not constitute a severe restriction. (See Fodor [1984], and [1985], p. 99.)

But on the strict construal of the conditions under which a psychological taxonomy may be relational, the apparent externalism of Marr's theory has to be regarded as a largely heuristic feature. Except where teleological explanations are being offered, the externalism is not to be taken with full seriousness. Since there is no indication in Marr's work that the externalism is anything other than an integral feature of the theory, Fodor's position is to some extent revisionary. His claim that a scientific psychology should not match the externalism of the commonsense scheme thus assumes considerable importance.

3. If we ignore teleological explanations in psychology, then Fodor's major argument about causal powers would have the consequence that a scientific psychology should classify together the neurophysiologically identical twins of Putnam's and Burge's examples. For Fodor's major argument about causal powers is designed to show that the psychological states of such twins do have the same causal powers.

The core of the argument is as follows. Science is in the business of causal explanation; causal explanation is a matter of subsuming events under causal generalisations; and subsuming events under causal generalisations involves classifying events by their causal powers. This is partly constitutive of the notion of a causal power.

All this can be agreed, along with the point that an event can be subsumed under just the same causal generalisations *of particle physics* whether the particles involved are H-particles or T-particles—where any particle is an H-particle if Fodor's dime is heads up, and is a T-particle if Fodor's dime is tails up (p. 241). What does this undisputed core show about classification in psychology?

Suppose that it has not yet been ruled out—as at the stage of the argument under review it has indeed not been ruled out—that the counterfactual supporting generalisations of psychology, like the workaday generalisations of the common-sense scheme, are intensional and externalist. The *prima facie* credentials of this thesis would seem, after all, to be very much on a par with the *prima facie* credentials of any propositional attitude psychology. Then the taxonomy that will facilitate subsumption under these generalisations will—without any special generosity—itself be externalist, for all that the core argument of the last paragraph shows. So, if that core argument had to bear the onus for unseating a presumption that psychology is externalist, then it would seem quite powerless to do so. However, in Fodor's hands the core argument does not bear that onus, for within his major argument about causal powers the core argument is augmented in various ways.

The friend of externalism in psychology is first imagined to maintain that the states which he classifies differently *do* have different causal consequences. The reply, augmenting the core argument, is that these are different consequences in different contexts or enviroments, and that the states still have the same causal *powers* since they would have the same consequences in the same context or environment (p. 244). However, this reply is dialectically unconvincing against a committed externalist, as Fodor himself notes (p. 246). The externalist holds that the content of a mental state is partly determined by external, environmental, factors. Consequently, it would be question

begging simply to insist on considering the same mental state, with the same content, in a different environment from the one in which it is actually set.

Furthermore, this reply—this augmentation of the core argument— is not obviously consistent with another element of Fodor's position. Quite apart from the issue of teleological explanation, it is part of Fodor's position that a scientific taxonomy may be relational. He gives as an example the relational property of being a planet, deployed in astronomy. There is a causal generalisation that connects being a planet and moving in an ellipse. Being a planet has actual, present, causal consequences for one's trajectory; being a planet affects—causally affects—one's causal powers (p. 250). Since science is in the business of subsuming events under causal generalisations, the property of being a planet can count in a scientific taxonomy.

It is this part of Fodor's position that seems to be in tension with the insistence that causal powers must be compared across contexts or environments. For it cannot be the case *both* that a planet has characteristic causal powers and not merely those of a physically similar chunk of matter that is not a planet, *and* that causal powers have to be compared across contexts or environments quite generally.

I take it that, roughly enough for present purposes, a chunk of matter, whether or not it is a planet, is subject to the inverse square law of gravitational attraction, and that if a chunk of matter is moving sufficiently fast in the environment of a much heavier body then its trajectory will be an ellipse; in fact, an ellipse around that body. But it is not true, concerning something that is in fact a planet moving around body *a*, say, that if it were in a different environment it would still move in an ellipse; even less that if it were in a different environment it would still move in an ellipse around body *a*.

The requirement that causal powers be compared across contexts or environments quite generally seems to have the consequence that the property of being a planet cannot count taxonomically. Since that consequence is evidently incorrect, the requirement should be dropped or refined. Astronomy is, I take it, like psychology in being a *special science*. (See Fodor [1974].) Its generalisations are not totally general. While they apply over a range of actual and possible contexts, they—and

the taxonomy they deploy—presume upon certain environmental factors not being varied. A planet is part of a solar system, for example, and astronomical generalisations may presume upon that relationship of embedding within a system not being varied.

If the environment of what is in fact a planet had been different in certain ways, then the generalisations of astronomy might have been different, or there might have been no worthwhile generalisations to be captured; the taxonomy of astronomy might have classified the same bodies in different ways, or might not have been applicable at all. For the notion of causal power that goes along with the taxonomy of a special science, it is not legitimate to require that causal powers be compared across contexts that fall outside the range of the generalisations.

If the requirement is dropped, then the property of being a planet is allowed to count taxonomically. So too, on the face of it, is the property of being a planet of body *a*. Whether you are a planet of *a* affects your trajectory—an ellipse around *a*—and your trajectory determines what you can bump into. As a planet of *a*, you are unlikely to bump into *a*, for example—at least in the near future. But the property of being a planet of *a* does not, of course, taxonomically *supplant* the property of being a planet *simpliciter*. The science of astronomy can use the coarser taxonomy as well as the more refined one. There is something of significance for astronomy that planets of *a* and planets of *b* have in common, namely, both move in ellipses; and that generalisation ought to be captured by the science.

Similarly, the most committed externalist about psychology can also allow for various coarser taxonomies that generalise across certain environmental differences, provided—what is not obviously guaranteed *a priori*—that there is something of psychological significance remaining once we abstract from those environmental factors. (Cf. Kitcher [1985], pp. 87-8.)

4. The first augmentation of the core argument is dialectically unconvincing. To avoid a stand off, Fodor augments the core argument in a second way. The core argument claimed that science classifies by causal powers, and that the classifications of particle physics are indifferent to the state of Fodor's dime. The

core argument is now augmented by the claim that if a scientific taxonomy is to be sensitive to an environmental feature then, since it is a taxonomy by causal powers, the causal powers of the classified items must depend upon that environmental feature; and the dependence must be mediated by causal mechanisms or causal laws (pp. 247–8). This augmentation is intended to provide a kind of *reductio ad absurdum* of the externalist who is trying to maintain that his taxonomy is still a scientific taxonomy. For in the Putnam and Burge examples there are no mechanisms or laws that connect environmental features and the psychological states of the twins.

However, we can have at least two reservations about this argument. First, what is supposed to be absurd about the idea that environmental differences *cause* differences between the twins is that one cannot cause differences in the causal powers of a person's psychological states without changing the person physiologically; yet the twins in the example are physiologically identical. But it is important to recall the dialectical situation. It is the presumption that psychology is externalist that has to be unseated. Suppose that psychological generalisations, and the corresponding causal powers, are conceived as the externalist conceives them. Then psychological states differing in that one is a state of looking at object *c* and the other is a state of looking at object *d* do differ in their causal powers. And the externalist can maintain this without denying the possibility of a coarser taxonomy as well. But such a difference can obviously be brought about without any change in the perceiver's physiology: just switch *c* and *d* without the subject noticing.

The second reservation is this. The argument seems to rest upon an assumption that if a causal power depends upon an environmental factor, then the dependence is a causal one. Now, a typical consequence of externalism is that, if a neurophysiological twin of an actual subject *had* been set in a different environment then our actual taxonomy would not have applied in the counterfactually imagined environment to classify the twin in the same way as the actual subject is classified. In that sense, he *would have been* psychologically different from the actual subject. But it does not follow from this that a way of *making* the actual subject psychologically different in those ways now is by *changing* his environment now. Still less does it follow from the externalist

claim about the counterfactual environment, that we can *make* the subject psychologically different now without making any physiological difference. So even if it were absurd to suppose that one can make a psychological difference without making a physiological difference, that would not constitute a *reductio* of externalism about scientific psychology. For externalism need not have that consequence.

To highlight the fact that a classificatory scheme is externalist, we imagine a highly counterfactual situation with respect to which the classificatory scheme yields a different classification because of an external difference. We do not thereby commit ourselves to the view that if practitioners of the science in question were set in a world in which pairs of cases differed in just that external respect, then the scientists would deploy that same classificatory scheme in their explanations. (On this point, see Burge [1986a], p. 21-2.) Nor do we commit ourselves to any general answer to the question how the classificatory scheme would need to differ from the actual scheme. Different examples may well require different answers.

5. I have been considering Fodor's major argument about causal powers. The task for that argument was to overcome a presumption that the generalisations of a scientific psychology are typically externalist. I have claimed that the core argument is incapable of that task. The first augmentation is dialectically unconvincing and seems to lead to a stand off. It is also in tension with other elements of Fodor's position, and insists on a requirement which is not legitimate in the special sciences. I have also expressed two reservations about the second augmentation, namely that what is supposed to be absurd has not been demonstrated to be absurd, and that what is supposed to be absurd has not been shown to be a consequence of the externalist's position.

Much more discussion is required. In particular, the various examples of externality need to be examined in detail, and the differences between the cases need to be articulated. Pending that kind of examination, I have argued that the claim that scientific psychology is, and will continue to be, externalist has not been shown to be untenable.

II

6. I turn now to Fodor's solution to the apparent problem posed by what he regards as the inevitable mismatch between the commonsense scheme and scientific psychology. The problem is that for any notion of content available to a scientific psychology content will not determine truth conditions. In my view, a helpful way to assess Fodor's solution is to consider the partially analogous problem posed for the semantics of natural language by the fact that sentence meaning does not typically determine truth conditions.

Putnam's original Twin Earth examples were, of course, presented as a problem for certain conceptions of linguistic meaning, and Burge's examples (in [1979]) are also closely related to language. But I shall consider the phenonemon of context dependence in a more general and schematic way.

To see what the analogue of Fodor's problem would be in this case, begin by imagining someone who accepts that the notion of sentence meaning is 'at least a little problematic' (p. 254). In partial elucidation of the notion this theorist might offer the claim that meaning at least determines truth conditions, so that differences of truth conditions—and *a fortiori* differences of truth value—require differences of meaning. He might even venture the suggestion that this is almost all that is clear about meaning. Now imagine this theorist to be reminded that, however things might be with formal languages, the linguistic meaning of a natural language sentence containing an indexical word—'I', 'here', 'now'—or a demonstrative—'that', 'this'— or complex demonstrative—'that blue carnation', 'this hedgehog'—does not, by itself, determine truth conditions. Two utterances of the very same sentence may have different truth conditions. Responses to this reminder might well vary from 'So what?' to something about a fly in the ointment. (See Davidson [1967], p. 33.) But it is unlikely that the facts about indexicals and demonstratives would immediately be perceived as presenting a problem for the very notion of sentence meaning. In order to turn those facts into a problem about meaning, one needs a *diagnosis*. It would go like this: Indexicals and demonstratives break the connection between meaning and truth conditions, and that connection was almost all that was clear about

meaning; without it we have no idea how to judge sameness and difference of meaning.

The obvious reply would be that indexicals and demonstratives do not *break* the connection between meaning and truth conditions; 'they just relativize it to context' (p. 255; and cf. Putnam [1975], p. 234). The evident correctness of this line of reply suggests that *if* the externality of the truth conditional content of psychological states is analogous to the context dependence of natural language, then Fodor's notion of narrow content is in no worse shape than the notion of sentence meaning. And that seems to be an optimistic thought.

There are many delicate issues here—some of them issues of interpretation. For example, at one point Putnam adopted a treatment of the Twin Earth examples in terms of 'an unnoticed indexical component' ([1975], p. 234), although that is not the only, or even the most important, aspect of his total view; and it is certainly not the aspect most congenial to those who see major similarities between Putnam and Burge. (See Burge [1979], Note 2, and Burge [1982a].) Also, Fodor, in an earlier paper, convincingly rejected an indexicality treatment of the examples. (See Fodor [1982], and cf. Burge [1982b].) But the version that he rejected was a metalinguistic one, and his reasons for rejection might not generalise to other versions.

I propose to turn my back on these delicacies in order to make a point about the final position that Fodor articulates in response to the Still Small Voice. There are two main aspects to that final position. One is that, because of a certain *inexpressibility* of narrow content, and because the term 'content' might be reserved for truth conditional content, the account of narrow content that emerges is 'in a certain sense, a "no content" account' (p. 261). The other is that, because narrow contents—while not themselves truth conditional—constitutively determine functions from contexts to truth conditions, the account of narrow content that emerges is 'a fully intensionalist account' (*ibid*). There are imaginable worries about that second aspect; but here I am going to focus on the first.

7. My claim is that the phenonenon of inexpressibility that Fodor discusses is not symptomatic of a 'no content' account, in any sense of that term. In a little more detail, inexpressibility can

arise within accounts which in no way prescind from content—from the semantic or representational properties of utterances and thoughts—and it can arise even for a notion of truth conditional content. I propose to illustrate the point by considering semantic theories for sentences containing indexicals and demonstratives, and neo-Fregean theories of content. (For the neo-Fregean approach, with its stress on non-descriptive modes of presentation, see for example Burge [1977], Evans [1981] and [1982], and Peacocke [1981] and [1983].)

It is a familiar Fregean claim that a full account of the content of propositional attitudes requires the notion of a mode of presentation, and neo-Fregeans have attempted to give substantive theories of modes of presentation (types), or ways of thinking of objects and properties. By no stretch of the imagination do neo-Fregean theories prescind from the semantic or representational properties of thoughts. Yet the phenonenon of inexpressibility is recognised by such theories.

Frege claimed that first person thoughts are incommunicable. (See Frege [1956], p. 26.) This is apt to sound mysterious, even though it is not. But, in any case, the phenomenon of inexpressibility is not restricted to first person thoughts. We can illustrate it with other indexicals. If someone rings me from Australia and tells me, 'It's 40°C here', then there is no thought that I can think that matches the speaker's thought both in which place is thought about and in the way that it is thought about—the 'here' way. In general, there is nothing that I can say that perfectly expresses the content of another man's 'here'-thoughts, unless I happen to be located at the same place as that other man. Similar remarks could be made about 'now'-thoughts.

Suppose that a man looks at a particular blue carnation and thinks, 'That blue carnation is pretty'. Then, if I cannot see the carnation in question then I cannot think a thought that matches the man's thought both in which object is thought about and in the mode of presentation—a perceptual mode—of the object. And, as in the case of indexicals, so in the case of perceptual demonstratives, there is nothing that I can say that perfectly expresses the content of the man's thought.

One area in which this kind of phenomenon has been recognised is the formal semantics of sentences containing

indexicals and demonstratives. In a theory of truth conditions—approximately in the style of Davidson—we might find something like:

> If speaker *s* demonstrates object *x* in his utterance *u* at *t* of 'That blue carnation is pretty' then
> *u* is true iff *x* is pretty.

If, on a particular occasion of utterance, the demonstrated object is a carnation called 'Fido' then we can infer from the theory that the utterance is true iff Fido is pretty. Now, there is a way in which a theory with this consequence can seem to fall short as a theory of *interpretation.* For there is a dimension of semantic similarity along which a theorist's utterance of 'Fido is pretty' does not match the original utterance of 'That blue carnation is pretty'. And likewise there is a dimension of content similarity of thoughts along which the theorist's thought will not in general match the speaker's thought.

None of this has the consequence that one cannot give a full and accurate report on the content of an utterance containing indexical or demonstrative expressions. But such a report will have to be partially indirect. Our speaker said that Fido is pretty—or said of Fido that it is pretty—and referred to Fido by using a complex demonstrative 'that . . .'. Likewise, the thought he expressed was about Fido, to the effect that it is pretty, and he thought about Fido in a specified perceptual demonstrative way. As Christopher Peacocke put the point: 'We need . . . to invoke the distinction between referring to and employing modes of presentation' ([1981], p. 192).

The lesson is that a thinker in one context may be unable to match the thought of a thinker in another context in point of both reference and mode of presentation. But this problem of 'inexpressibility'—which arises for a notion of truth conditional content—can be overcome by referring to modes of presentation, perhaps via linguistic expressions conventionally associated with them.

Within a neo-Fregean framework, this strategy of referring to modes of presentation can be used to overcome a slightly different problem of 'inexpressibility'. Two thinkers may think of different objects in a similar way, and it may be important to generalise over thoughts that differ in reference, yet are similar

in point of mode of presentation. But there is no thought which has as its content precisely what the similar thoughts have in common, and no more. Similarly, in semantic theories we generalise over utterances that contain the same indexical expression. But there is nothing that one can say that matches the utterances along the dimension of semantic similarity, and also does not differ from any of those similar utterances in any semantic respect. The way to specify what is in common between the utterances is to mention, rather than use, the indexical expression and specify its semantic properties—the properties that determine reference in context. Similarly, the way to specify what is in common between similar thoughts is to refer to, rather than to employ, the common mode of presentation.

These latter claims about what various indexical utterances may have in common, and about what various indexical thoughts may have in common, are, as Fodor himself notes, analogous to what he wants to claim about narrow content: 'You can't, in absolute strictness, *express* narrow content; but . . . there are ways of sneaking up on it' (p. 261). It is clear that there is no interesting sense in which a 'no content' account could emerge from the claims about indexicals; and surely the same goes for the inexpressibility of narrow content. (A 'no content' account—on the natural construal of the term—would be a purely syntactic account in the spirit of Stich [1983].)

There is just one sense in which the account of narrow content has been shown to be a 'no content' account: narrow content does not by itself determine truth conditions. But the phenomena of inexpressibility are not intrinsically connected with the failure to determine truth conditions. That was illustrated in the earlier examples of the impossibility of one person perfectly matching the indexical and demonstrative complete thoughts of another.

Whether or not a scientific psychology actually requires the notion of narrow content, this point about Fodor's response to the Still Small Voice is not one that endangers that notion. The suggestion is, rather, that in the first aspect of his final position Fodor concedes more to the Voice than he needs to.

8. In the discussion of inexpressibility, I have used the neo-

Fregean position as an analogy. There are, however, important differences between the neo-Fregean notion of a mode of presentation (type) or way of thinking and Fodor's notion of narrow content.

To be a neo-Fregean is not, of course, to say that ways of thinking or modes of presentation have a life of their own, and are only accidentally or incidentally modes of presentation *of objects*. On the contrary, modes of presentation (types) are best conceived as abstractions over presentations of particular objects—mode of presentation tokens—which are similar along a certain dimension. These latter are constituents of complete thoughts—thoughts that are evaluable for truth or falsehood. What is being abstracted from is a certain environmental factor, namely *which* object is being presented. This conception of modes of presentation is independent of the claim, made by some neo-Fregeans, that demonstrative thoughts are *existence dependent*—the claim, that is, that if a demonstrative thought concerns a particular object then the thinker could not entertain that thought in the absence of that object. (See, for example, Evans [1982].)

Now, it is consistent to maintain—and arguably correct—that, although a mode of presentation determines a (partial) function from contexts to objects of thought, it is still the case that the notion of a mode of presentation and its associated function is conceptually posterior to the notions of a complete thought and its constituents—that the constitutive account of a mode of presentation will advert to the relations between thinkers and the objects typically presented under that mode.

Earlier in his paper, Fodor appears to deny the analogous claim of conceptual posteriority about narrow content, when he says:

> The 'broad content' of a thought . . . is . . . what you get when you specify a narrow content and fix a content. *This makes* the notion of narrow content the more basic of the two; which is just what sensible people have always supposed it to be. (p. 257; my italics)

However, it is not so clear that the claim that narrow content is the posterior notion is being denied in the latter part of the paper where we have:

> Narrow content is . . . only content *potentially*; it's what gets to *be* content when—and only when—it gets to be anchored. (p. 259)

This looks much more congenial to the view that narrow content is constitutively the result of generalising over contents, abstracting from environmental anchoring.

A neo-Fregean makes use of three notions of content. Comparing contents *simpliciter* (or *extrinsic* contents) along one dimension of similarity—similarity of reference, whatever the mode of presentation—yields a notion of *referential* content. Comparing them along another dimension—similarity of mode of presentation, whatever the reference—yields a notion of *unanchored* content. In order to highlight a point about inexpressibility, I have been drawing an analogy between the neo-Fregean's unanchored content and Fodor's narrow content. However, if narrow content is governed by a principle of supervenience upon the neurophysiological, then the two notions should not be identified.

Consider again the case of perceptual demonstrative thoughts. A thought that a particular blue carnation, Fido, is pretty could have its (extrinsic) content represented:

$$\langle\langle \text{Fido}, \langle P, \textit{being a blue carnation}\rangle\rangle, \textit{being pretty}\rangle.$$

(The idea behind this scheme of representation is simply that it specifies which object was thought about, the mode of presentation of that object, and what was thought about that object. The element '*P*' stands in for the perceptual character of the experience in which Fido is presented.) The unanchored content—which abstracts from the fact that it is Fido that is thought about—could then be represented:

$$\langle\langle P, \textit{being a blue carnation}\rangle, \textit{being pretty}\rangle.$$

Now, leaving aside the concept *being pretty*, which has a predicative role in the thought, it is still the case that the mode of presentation of Fido, $\langle P, \textit{being a blue carnation}\rangle$, involves the concept *being a blue carnation.* What is more, the involvement of a concept in an individuative—rather than predicative—role is crucial if it is to be the case that mode of presentation plus context determines reference. (See Peacocke [1981], p. 201, and

my [1982].) But typical externalist points can be made about this concept.

In the case under consideration, Fido is presented to our thinker in a visual experience *as of* a blue carnation. But the 'as of' character of a visual experience does not supervene upon the neurophysiology of the perceiver. Consequently, neo-Fregean modes of presentation do not, generally, respect supervenience.

In fact, Burge gives a general argument against individualism —in *his* sense, of course—concerning the contents of visual experiences. (See Burge [1986a], Section 3, and [1986b].) The example given is structurally analogous to the examples in 'Individualism and the Mental', but it depends on neither linguistic nor social factors. The main idea is simply that the attribution of content to an experience answers in part to what condition in the world normally produces that kind of experience.

In sum, then, the neo-Fregean notions of unanchored content and mode of presentation are conceptually derivative notions, while it is not so clear whether the notion of narrow content is posterior to the notion of broad content. And the neo-Fregean notions do not respect supervenience, whereas narrow content does.

* * *

9. I have argued in Part I that the externality of the commonsense scheme does not, by itself, constitute a problem for a content based scientific psychology. In Part II, I have claimed that, if we are to construct a notion of narrow content, then the phenomenon of inexpressibility does not constitute any kind of obstacle. I have not dissented from Fodor's overall conclusion that the prospects for an attitude based scientific psychology are no worse than before the Putnam and Burge examples were introduced.

However, there are several lines of argument that have not been explored at all in this paper. For example, first, what I have argued in the first part of the paper leaves it open that particular features of the externality of the commonsense scheme may unsuit the attitudes attributed in the scheme for incorporation into a scientific psychology. Second, the context dependence of

natural language is not generally assimilable to indexicality, and this fact has seemed to some to threaten the idea of sentence meaning. Similarly, there is a more radical way of viewing the externality of the commonsense scheme; and on that view there is no reason to suppose that any notion of content can respect supervenience. (See especially the Editors' Introduction to McDowell and Pettit [1986].) Third, in respect of externality and supervenience, there may be important differences between the psychology of input systems or *modules* and the psychology of the central cognitive system. (For the distinction, and for pessimism about the prospects for a scientific psychology of central systems, see Fodor [1983].) Finally, it might be that there is a significant gap between content based psychology in general and attitude based psychology in particular, and that threats to the latter are not *ipso facto* threats to the former.

Principled confidence about an attitude based scientific psychology must wait upon detailed exploration of all these lines of argument.[1]

[1] For comments on earlier versions of this paper, I am grateful to Dorothy Edgington, Samuel Guttenplan, Ian McFetridge, Andrew Woodfield, and especially, Christopher Peacocke.

REFERENCES

Burge, T.
1977 'Belief *De Re*', *Journal of Philosophy* vol.74 (1977), pp. 338-62
1979 'Individualism and the Mental', in P. A. French *et al* (eds.), *Midwest Studies in Philosophy Volume 4: Studies in Metaphysics* (Minneapolis: University of Minnesota Press, 1979), pp. 73-121
1982a 'Other Bodies', in A. Woodfield (ed.), *Thought and Object* (Oxford University Press, 1982), pp. 97-120
1982b 'Two Thought Experiments Reviewed', *Notre Dame Journal of Formal Logic* vol.23 (1982), pp. 284-93
1986a 'Individualism and Psychology', *Philosophical Review* vol.95 (1986), pp. 3-45
1986b 'Cartesian Error and the Objectivity of Perception', in J. McDowell and P. Pettit [1986]

Davidson, D.
1967 'Truth and Meaning', *Synthese* vol.17 (1967), pp. 304-23; reprinted in Davidson [1984], pp. 17-36 (Page reference to reprinting.)
1984 *Inquiries into Truth and Interpretation* (Oxford University Press, 1984)

Davies, M.
1982 'Individuation and the Semantics of Demonstratives', *Journal of Philosophical Logic* vol.11 (1982), pp. 287-310
1983 'Function in Perception', *Australasian Journal of Philosophy* vol.61 (1983), pp. 409-26

Evans, G.
1981 'Understanding Demonstratives', in H. Parret and J. Bouveresse (eds.), *Meaning and Understanding* (Berlin: De Gruyter, 1981), pp. 280-303; reprinted in Evans [1986], pp. 291-321
1982 *The Varieties of Reference* (Oxford University Press, 1982)
1986 *Collected Papers* (Oxford University Press, 1986)

Fodor, J. A.
1974 'Special Sciences', *Synthese* vol.28 (1974), pp. 77-115; reprinted in Fodor [1981], pp. 127-45
1980 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology', *Behavioral and Brain Sciences* vol.3 (1980), pp. 63-73; reprinted in Fodor [1981], pp. 225-253
1981 *Representations* (Brighton: Harvester Press, 1981)
1982 'Cognitive Science and the Twin-Earth Problem', *Notre Dame Journal of Formal Logic* vol.23 (1982), pp. 98-118
1983 *The Modularity of Mind* (Cambridge: MIT Press, 1983)
1984 'Semantics, Wisconsin Style', *Synthese* vol.59 (1984), pp. 231-50
1985 'Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum', *Mind* vol.94 (1985), pp. 76-100

Frege, G.
1956 'The Thought: A Logical Inquiry', in P. F. Strawson (ed), *Philosophical Logic* (Oxford Univeristy Press, 1967), pp. 17-38

Kitcher, P.
1985 'Narrow Taxonomy and Wide Functionalism', *Philosophy of Science*, vol.52 (1985), pp. 78-97

McDowell, J. and Pettit, P.
1986 *Subject, Thought, and Context* (Oxford University Press, 1986)

Marr, D.
1982 *Vision* (San Francisco: W. H. Freeman and Company, 1982)

Peacocke, C.
1981 'Demonstrative Thought and Psychological Explanation', *Synthese* vol.49 (1981), pp. 187-217
1983 *Sense and Content* (Oxford University Press, 1983)

Putnam, H.
1975 'The Meaning of "Meaning" ', in K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science Volume 7* (Minneapolis: University of Minnesota Press, 1975), pp. 131-93; reprinted in *Philosophical Papers Volume 2* (Cambridge University Press, 1975), pp. 215-71 (Page references to reprinting.)

Stitch, S.
1983 *From Folk Psychology to Cognitive Science: The Case Against Belief* (Cambridge: MIT Press, 1983)