belief/desire explanations are vindicated if scientific psychology is ontologically committed to beliefs and desires. But it's *not* also required that the folk-psychological inventory of propositional attitudes should turn out to exhaust a natural kind. It would be astounding if it did; how could common sense know all that? What's important about RTM—what makes RTM a vindication of intuitive belief/desire psychology—isn't that it picks out a kind that is precisely coextensive with the propositional attitudes. It's that RTM shows how intentional states could have causal powers; precisely the aspect of commonsense intentional realism that seemed most perplexing from a metaphysical point of view.

Molecular physics vindicates the intuitive taxonomy of middle-sized objects into liquids and solids. But the nearest kind to the liquids that molecular physics acknowledges includes some of what common sense would not; glass, for example. So what?

So much for RTM; so much for this chapter, too. There is a strong prima facie case for commonsense belief/desire explanation. Common sense would be vindicated if some good theory of the mind proved to be committed to entities which—like the attitudes—are both semantically evaluable and etiologically involved. RTM looks like being a good theory of the mind that is so committed; so if RTM is true, common sense is vindicated. It goes without saying that RTM needs to make an empirical case; we need good accounts, independently confirmed, of mental processes as causal sequences of transformations of mental representations. Modern cognitive psychology is devoted, practically in its entirety, to devising and confirming such accounts. For present purposes, I shall take all that as read. What the rest of this book is about is doubts about RTM that turn on its *semantic* assumptions. This is home ground for philosophers, and increasingly the natives are restless.

2

Individualism and Supervenience

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. Having once told you the stories, I will then spend most of this chapter trying to puzzle out what, if anything, they have to do either with commonsense belief/desire explanation or with RTM. The conclusion will be: not much. That may sound pretty dreary, but I've been to 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 lot of recent philosophical discussion. Indeed, contrary to the conclusion that I am driving toward, 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 from what 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 & Wagnalls *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: 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* toward 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 toward 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 the 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 will assume that what's at issue in the Putnam and Burge stories is 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, my Twin's water2-beliefs are supposed to differ in content from my water-beliefs, and what's supposed to account for the difference is the chemical composition of the stuff *in our respective environments*. Analogously, Oscar's brisket-beliefs are supposed to differ in content from Oscar2's brisket2-beliefs, and what's supposed to account for the difference is what the form of words 'is brisket' applies to *in their respective language communities*.)

Brian Loar, in a recent, important paper (*SCPC*), has argued that these concessions may be too generous. Loar points out that the standard interpretation of the Twin cases takes for granted that if, for example, the predicate 'believes that water is . . .' applies to me but not to my Twin, and the predicate 'believes that water2 is . . .' applies to my Twin but not to me, then it follows that the content of my belief differs in some respect from the content of my Twin's. In effect, according to Loar, Putnam and Burge assume that you can infer identities and differences in beliefs from corresponding identities and differences in the 'that . . .' clauses that are used to ascribe them; and Loar gives grounds for doubting that such inferences are invariably sound. I think Loar may well be right about this, but I propose to ignore it. It's interesting to see what would follow from assuming that people situated the way that the Twins and the Oscars are ipso facto believe different things, whether or not the Burge/Putnam intuitions actually show that they do.

In aid of which, I shall talk as follows: 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 'nonrelational.' 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 matter. 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 common sense understands them, don't supervene on brain states. To put it roughly: States of type *X* supervene on states of 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. But it's arguable that any scientifically useful notion of psychological state ought to respect supervenience; mind/brain supervenience (and/or mind/brain identity) is, after all, the best idea that anyone has had so far about how mental causation is possible. The moral would appear to be that you can't make respectable science out of the attitudes as commonsensically individuated.

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, other identity conditions for mental states than those that common sense 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 such relational properties as, for example, *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 brain states are individuated relationally, then they will themselves fail to supervene 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 to quantum mechanics?")

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 notion of content actually can be constructed. Which they do. Vehemently.

This standard understanding of the difference between the way that common sense construes the attitudes and the way that psychology does is summarized as follows:

Commonsense Taxonomy (Pattern A)

1. Individuates the attitudes relationally; hence, assumes a nonindividualistic notion of content.
2. Distinguishes: my beliefs from my Twin's,
   Oscar's beliefs from Oscar2's.
3. Individuates brain states nonrelationally; therefore:
4. Violates supervenience.[1]

Psychological Taxonomy (Pattern B)

1. Individuates the attitudes nonrelationally; hence, assume a narrow notion of content.
2. Identifies: my beliefs with my Twin's,
   Oscar's beliefs with Oscar2's.
3. Individuates brain states nonrelationally; therefore:
4. Preserves supervenience.

One can imagine quite a different reaction to the Twin examples, however. According to this revisionist account, psychology taxonomizes the attitudes precisely the same way that common sense 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;[2] 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 common sense 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 that militate for the nonrelational individuation of mental states (hence, for preserving supervenience at the cost of violating the commonsense taxonomy) are no different from the ones that militate for the nonrelational individuation of brain states, molecular states, and such. This becomes evident as soon as one understands the source of our commitment to nonrelational taxonomy in these latter cases.

All this takes some proving. I propose to proceed as follows: First, we'll consider why we think that brain states and the like 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 scientific 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 then be back exactly where we started: Common sense 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 scientifically defensible intentional psychology 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*. By facing my dime heads-up, I now bring it about that every particle in the universe is an *H*-particle . . . thus! And then, by reversing my dime, 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* the universe, changing *H*–brain states into *T*–brain states 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 that 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 cetain 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.[3] 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.

In short, 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 see why it would be mad to embrace a taxonomy that 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 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, mutatis mutandis, for the properties of being *H* and *T* *brain states*. And similarly, mutatis mutandis, for the properties of being *H* and *T* *mental states*. *And similarly, mutatis mutandis, 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 individualistically. And 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 common sense deploys a taxonomy that *does* distinguish between the mental states of Oscar and Oscar2. So the commonsense taxonomy won't do for the purposes of psychology. Q.E.D.[4]

I can, however, 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, 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 lift it. 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 *x*'s *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 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 is the following pair of counterfactuals: (a) If his 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.

This line of argument no doubt assumes that I *can* say "Bring water!" on Twin-Earth—that my being on Twin-Earth doesn't ipso facto change my dialect to English2 (and, mutatis mutandis, convert my concept *water* into the concept *water2*). But although I've heard it

suggested that mental states construed nonindividualistically are easily bruised and don't 'travel,' the contrary assumption would in fact seem to be secure. The standard intuition about 'visiting' cases is that if, standing on Twin-Earth, I say "That's water" about a puddle of XYZ, then what I say is *false*. Which it wouldn't be if I were speaking English2.

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. On the contrary, since the causal subjunctives about the two states are the same, it must be that they have the *same* causal powers and thus count as the same state by what we're taking to be the relevant typological criteria.

Second line: "Maybe the causal powers of the mental states of Twins are always the same when their effects are *non*intentionally individuated. But consider their effects as intentionally described; consider, in particular, 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, for example, a meat counter), behave brisket2-purchasingly.[5] And, of course, when he says and does these things with brisket2 in mind, he may produce precisely the same bodily *motions* as his counterpart produces when he says and does the corresponding things with brisket in mind. 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's 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 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 effects *are* identical; 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. (Burge—who would, of course, accept this conclusion on independent grounds—nevertheless objects that the present sort of argument misunderstands the function of his and Putnam's thought experiments: Since the examples concern the description of circumstances presumed to be counterfactual, the likelihood or otherwise of such circumstances *actually occurring* is not, according to Burge, a relevant consideration. (See *IP*.) But this misses a point of methodology. We do, of course, want to tell the right story about how counterfactual circumstances should be described qua counterfactual. But we *also* want to tell the right story about how such circumstances should be described if they were real. The present intuition is that, were we actually to encounter Twins, what we should want to say of them is *not* that their quite different mental states have somehow managed to converge on the same behaviors; we *can* imagine examples that we'd want to describe that way, but Twins aren't among them. Rather, what we'd want to say about Twins is just that the (putative) differences between their minds are reflected, in the usual way, by corresponding differences between their behaviors. But we *can* say this only if we *are* prepared to describe their behaviors as different. So again it turns out that anyone who counts in a way that distinguishes the minds of Twins should also count in a way that distinguishes their acts.)

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. Claiming, 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.

First reply: If this argument shows that my mental state differs from my Twin's, it's hard to see why it doesn't show that our brain states differ too. My Twin is in a brain state that eventuates in his uttering the form of words 'Bring water.' I am in a brain state that eventuates in my uttering the form of words 'Bring water.' If our uttering these forms of words counts as our behaving differently, then it looks as though our brain states differ in their behavioral consequences, hence in their causal powers, hence in the state types of which they are tokens. (Similarly, mutatis mutandis, for our quantum mechanical states.) But I thought we agreed a while back that it would be grotes-

que to suppose that brain states that live on Twin-Earth are ipso facto typologically distinct from brain states that live around here.

Second reply: Notice that corresponding to the present argument for a taxonomic distinction between my mental state and my Twin's, there is the analogous argument for distinguishing *H*-particles from *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, misled 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 that *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 brain states relationally. Moreover, it is very plausible that all these intuitions 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 Centauri, 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 Centauri 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.[6] If every instance of brisket-chewing behavior ipso facto counts as an event distinct in kind 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 that connects the causal powers of Oscar's mental states with the character of the speech community he lives in *and that 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,

of course. It's a contingent fact about how God made the world. God made the world such that the mechanisms by which environmental variables affect organic behaviors run via their effects on the organism's nervous system. Or so, at least, all the physiologists I know assure me.

Well then, in order to avoid postulating crazy causal mechanisms, we shall 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 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 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 percent of the variance is genetic and the remaining 72.7 percent is the result of early training. If, however, she counts Oscar and Oscar2 in, and if she counts their food preferences the way 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?[7]

I think it's perfectly clear how Psyche ought 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 states 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.

So, then, to bring this all together: You can affect the relational properties of things in all sorts of ways—including by stipulation. But for one thing to affect the causal powers of another, there must be a mediating law or mechanism. It's a mystery what this could be in the Twin (or Oscar) cases; not surprisingly, since it's surely plausible that the only mechanisms that *can* mediate environmental effects on the causal powers of mental states are neurological. The way to avoid making this mystery is to count the mental states—and, mutatis mutandis, the behaviors—of Twins (Oscars) as having the same causal powers, hence as taxonomically identical.

So much for the main line of the argument for individualism. Now just a word to bring the reader up to date on the literature.

In a recent paper (*IP*), Burge says that reasoning of the sort I've been pursuing "is confused. The confusion is abetted by careless use of the term 'affect,' conflating causation with individuation. Variations in the environment that do not vary the impacts that causally 'affect' the subject's body may 'affect' the individuation of the . . . intentional processes he or she is undergoing. . . . It does not follow that the environment causally affects the subject in any way that circumvents its having effects on the subject's body" (*IP*, p. 16). But it looks to me like that's precisely what *does* follow, assuming that by "causally affecting" the subject Burge means to include determining the causal powers of the subject's psychological states. You can't both individuate behaviors Burge's way (viz., *non*locally) and hold that the causal powers of mental states are locally supervenient. When individuation is *by* causal powers, questions of individuation and causation don't divide in the way that Burge wants them to.

Consider the case where my Twin and I both spy some water (viz., some H_2O). My seeing the stuff causes me to say (correctly) "That's water!" His seeing the stuff causes him to say (incorrectly) "That's water2!" (His saying this sounds just like my saying "That's water!" of course.) These sayings count as *different behaviors* when you individuate behaviors Burge's way; so the behavioral effects of seeing water are different for the two of us; so the causal powers of the state of seeing water are different depending on which of us is in it. And this difference is uniquely attributable to differences in the contextual background; aside from the contextual background, my Twin and I are identical for present purposes. So if you individuate behavior Burge's way, differences in contextual background effect differences in the causal powers of mental states without having correspondingly different "effects on the subject's body"; specifically, on his neural structure. But is Burge seriously prepared to give up the local supervenience of causal powers? *How could* differences of context affect the

causal powers of one's mental states without affecting the states of one's brain?

Burge can say, if he likes, that mind/brain supervenience be damned; though, as I keep pointing out, if mind/brain supervenience goes, the intelligibility of mental causation goes with it. Or he can save mind/brain supervenience by going contextual on *neurological* individuation. (As, indeed, he appears to be tempted to do; see his footnote 18 in *IP*. Here both intuition and scientific practice clearly run against him, however.) But what he can't do is split the difference. If supervenience be damned for individuation, it can't be saved for causation. Burge says that "local causation does not make more plausible local individuation" (p. 16), but he's wrong if, as it would seem, "local causation" implies local supervenience of causal powers. Local causation *requires* local individuation when so construed. You can have contextual individuation if you insist on it. But you can't have it for free. Etiology suffers.

Well, if all this is as patent as I'm making it out to be, how could anyone ever have supposed that the standards of individuation appropriate to the psychologist's purposes are other than individualistic? I cast no aspersions, but I have a dark suspicion; I think people get confused 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*.[8]

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 nonrelational, 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 methodological 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 affect 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 causal powers, which is all the 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—or failing to have it—has no effect on causal powers.[9]

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 defense 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 to be 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, *IVM*, 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.

I repeat the main points in a spirit of recapitulation. There are two of them; one is about the methodology of science, and one is about its metaphysics.

Methodological point: Categorization in science is characteristically taxonomy by causal powers. Identity of causal powers is identity of causal consequences across nomologically possible contexts.

Metaphysical point: Causal powers supervene on local microstructure. In the psychological case, they supervene on local neural structure. We abandon this principle at our peril; mind/brain supervenience (/identity) is our only plausible account of how mental states could have the causal powers that they do have. On the other hand, given what causal *powers* are, preserving the principle constrains the way that we individuate causal *consequences.* In the case of the behavioral consequences of the attitudes, it requires us to individuate them in ways that violate the commonsense taxonomy. So be it.

Well, I've gotten us where I promised to: back to where we started. There is a difference between the way psychology individuates behaviors and mental states and the way common sense does. At least there is if you assume that the Burge/Putnam intuitions are reliable.[10] 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 lakes; 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 Centauri 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 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). All right 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's 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 (that is, 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 difference of extension implies difference of content; so unclear are we now about what content-identity comes 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 status 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 satisfied in *W*. Thus, 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 all right for you to choose some other; my point is going to be that if you liked any story of even remotely 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 all 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, it's presumably common ground that there's 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 virtue 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 reemphasizing 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. 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*.

We can now see why we ought to reject both of the following two suggestions found in Putnam, *MM*: 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. (Since it's untendentious that stereotypes supervene on physiology, the stereotypes for real water and Twin-water must be identical; so if stereotypes did fix extensions, my Twin's 'water'-thoughts would have the same extension as mine.) 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."

Oh, good! Do proceed.

"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 intensions, 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; it's not the construct but the question that is 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., 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 'Griceian' theory of meaning. According to Griceian 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 got a sore toe' runs through your head and also runs through mine; what's the content that these thoughts share? Well, you can't say what it is by using the sentence "I've got a sore toe," since, whenever you use that sentence, the "I" automatically gets anchored to you. You can, however, sneak up on the shared content by *mentioning* that 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.[11] People who positively *insist* on asking them generally get what they deserve: phenomenalism, 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 thought that you share with your Twin?"

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 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 world 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 sorts of connections you imagine 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 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 a kind of vat too.

SSV: "But if a brain is a function from contexts to truth conditions, and if a vat can be a context, then when a brain in a vat thinks 'water is wet' the truth condition of its thought will be (not something about H_2O or XYZ but) something about its vat. So it will be thinking something *true*. Which violates the intuiton that the thoughts of brains in vats have to be *false* thoughts."

Answer: You're confused about your intuitions. What they really tell you isn't that the thoughts of brains in vats have to be false; it's that being in a vat wouldn't stop a brain from having the very thoughts that you have now. And that intuition is *true*, so long as you individuate thoughts narrowly. It's tempting to infer that if a brain has your thoughts, and has them under conditions that would make your thoughts false, then the thoughts that the brain is having must be false too. But to argue this way is exactly to equivocate between the narrow way of individuating thoughts and the broad way.

SSV: "Mental states are supposed to cause behavior. How can a function cause anything?"

Answer: Some functions are implemented in brains; and brains cause things. You can think of a narrow mental state as determining an equivalence class of mechanisms, where the criterion for being in the class is *semantic*.

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

Answer: 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 intentional 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. (See Stich, *FFPCS*.)

Now, sleep conscience!

What I hope this chapter 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 provide no such evidence—or that it constrains content attributions any less well than the old, unrelativized account used to do. The point to bear in mind is that if 'difference in extension → difference in intension' substantively constrains the attribution of propositional attitudes, then so too does this same principle when it is relativized to context. *The Moral:* If the worry about propositional attitudes is that Twin-Earth shows that contents don't determine extensions, the right thing to do is to *stop worrying*.

So it looks as though everything is all right. Super; Let, you might suppose, rejoicing be unconstrained. But if you do suppose that, that's only because you've let the Twin problems distract you from the hard problems. The hard problems start in chapter 3.[12]