13 Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum

Jerry A. Fodor

It rained for weeks and we were all *so* tired of ontology, but there didn't seem to be much else to do. Some of the children started to sulk and pull the cat's tail. It was going to be an *awful* afternoon until Uncle Wilifred thought of Mental Representations (which was a game that we hadn't played for *years*) and everybody got *very* excited and we jumped up and down and waved our hands and all talked at once and had a perfectly *lovely* romp. But Auntie said that she couldn't stand the noise and there would be tears before bedtime if we didn't please calm down.

Auntie rather disapproves of what is going on in the Playroom, and you can't entirely blame her. Ten or 15 years of philosophical discussion of mental representation has produced a considerable appearance of disorder. Every conceivable position seems to have been occupied, along with some whose conceivability it is permissible to doubt. And every view that anyone has mooted, someone else has undertaken to refute. This does *not* strike Auntie as constructive play. She sighs for the days when well-brought-up philosophers of mind kept themselves occupied for hours on end analyzing their behavioral dispositions.

But the chaotic appearances are actually misleading. A rather surprising amount of agreement has emerged, if not about who's winning, at least about how the game has to be played. In fact, everybody involved concurs, pretty much, on what the options are. They differ in their hunches about which of the options it would be profitable to exercise. The resulting noise is of these intuitions clashing. In this paper, I want to make as much of the consensus as I can explicit; both by way of reassuring Auntie and in order to provide new participants with a quick guide to the game: Who's where and how did they get there? Since it's very nearly true that you can locate all the players by their answers to quite a small number of diagnostic questions, I shall organize the discussion along those lines. What follows is a short projective test of the sort that self-absorbed persons use to reveal their hitherto unrecognized proclivities. I hope for a great success in California.

Reprinted from MIND Volume 94, 55–97 (1985) by permission of Oxford University Press and the author.

First Question: How Do You Feel about Propositional Attitudes?

The contemporary discussion about mental representation is intimately and intricately involved with the question of Realism about propositional attitudes. Since a goal of this essay is to locate the issues about mental representation with respect to other questions in the philosophy of mind, we commence by setting out this relation in several of its aspects.

The natural home of the propositional attitudes is in "commonsense" (or "belief/desire") psychological explanation. If you ask the Man on the Clapham Omnibus what precisely he is doing there, he will tell you a story along the following lines: "I wanted to get home (to work, to Auntie's) and I have reason to believe that there—or somewhere near it—is where this omnibus is going." It is, in short, untendentious that people regularly account for their voluntary behavior by citing beliefs and desires that they entertain; and that, if their behavior is challenged, they regularly defend it by maintaining the rationality of the beliefs ("Because it *says* it's going to Clapham") and the probity of the desires ("Because it's *nice* visiting Auntie"). That, however, is probably as far as the Clapham Omnibus will take us. What comes next is a philosophical gloss—and, eventually, a philosophical theory.

First Philosophical Gloss

When the ordinary chap says that he's doing what he is because he has the beliefs and desires that he does, it is reasonable to read the 'because' as a *causal* 'because'—whatever, exactly, a causal 'because' may be. At a minimum, common sense seems to require belief/desire explanations to support counterfactuals in ways that are familiar in causal explanation at large: If, for example, it is true that Psmith did A because he believed B and desired C, then it must be that Psmith would *not* have done A if either he had not believed B or he had not desired C. (Ceteris paribus, it goes without saying.) Common sense also probably takes it that if Psmith did A because he believed B and desired C, then—ceteris paribus again—believing B and desiring C is causally sufficient for doing A. (However, common sense does get confused about this since—though believing B and desiring C was what caused Psmith to do A—still it is common sense that Psmith could have believed B and desired C and *not* done A had he so decided. It is a question of some interest whether common sense can have it both ways.) Anyhow, to a first approximation the commonsense view is that there is mental causation, and that mental causes are subsumed by counterfactual-supporting generalizations of which the practical syllogism is perhaps the paradigm.

Closely connected is the following: Everyman's view seems to be that propositional attitudes cause (not only behavior but also) other propositional attitudes. Thoughts cause desires (so that thinking about visiting

Auntie makes one want to) and—perhaps a little more tendentiously—the other way around as well (so that the wish is often father to the thought, according to the commonsense view of mental genealogy). In the paradigm mental process—viz. thinking—thoughts give rise to one another and eventuate in the fixation of beliefs. That is what Sherlock Holmes was supposed to be so good at.

Second Philosophical Gloss

Common sense has it that beliefs and desires are semantically evaluable; that they have *satisfaction-conditions*. Roughly, the satisfaction-condition for a belief is the state of affairs in virtue of which that belief is true or false and the satisfaction-condition for a desire is the state of affairs in virtue of which that desire is fulfilled or frustrated. Thus, 'that it continues to rain' makes true the belief that it is raining and frustrates the desire that the rain should stop. This could stand a lot more sharpening, but it will do for the purposes at hand.

It will have occurred to the reader that there are other ways of glossing commonsense belief/desire psychology. And that, even if this way of glossing it is right, commonsense belief/desire psychology may be in need of emendation. Or cancellation. Quite so, but my purpose isn't to defend or criticize; I just want to establish a point of reference. I propose to say that someone is a *Realist* about propositional attitudes if (a) he holds that there are mental states whose occurrences and interactions cause behavior and do so, moreover, in ways that respect (at least to an approximation) the generalizations of commonsense belief/desire psychology; and (b) he holds that these same causally efficacious mental states are also semantically evaluable.

So much for commonsense psychological explanation. The connection with our topic is this: the full-blown Representational Theory of Mind (hereinafter RTM, about which a great deal presently) purports to explain how there *could be* states that have the semantical and causal properties that propositional attitudes are commonsensically supposed to have. In effect, RTM proposes an account of what the propositional attitudes *are*. So, the further you are from Realism about propositional attitudes, the dimmer the view of RTM that you are likely to take.

Quite a lot of the philosophical discussion that's relevant to RTM, therefore, concerns the status and prospects of commonsense intentional psychology. More, perhaps, than is generally realized. For example, we'll see presently that some of the philosophical worries about RTM derive from scepticism about the semantical properties of mental representations. Putnam, in particular, has been explicit in questioning whether coherent sense could be made of such properties. (See Putnam 1983, 1986.) I have my doubts about the seriousness of these worries (see Fodor 1985); but the present point is that they are, in any event, misdirected as arguments against RTM. If there is something wrong with meaning, what that shows is something *very* radical, viz. that there

is something wrong with propositional attitudes (a moral, by the way, that Quine, Davidson, and Stich, among others, have drawn explicitly). That, and *not* RTM, is surely the ground on which this action should be fought.

If, in short, you think that common sense is just plain *wrong* about the aetiology of behavior—i.e., that there is *nothing* that has the causal and semantic properties that common sense attributes to the attitudes—then the questions that RTM purports to answer don't so much as arise for you. You won't care much what the attitudes are if you take the view that there aren't any. Many philosophers do take this view and are thus united in their indifference to RTM. Among these Anti-Realists there are, however, interesting differences in motivation and tone of voice. Here, then, are some ways of not being a Realist about beliefs and desires.

First Anti-Realist Option

You could take an *instrumentalist* view of intentional explanation. You could hold that though there are, *strictly speaking*, no such things as belief and desires, still talking as though there were some often leads to confirmed behavioral predictions. Everyman is therefore licensed to talk that way—to adopt, as one says, the intentional stance—so long as he doesn't take the ontological commitments of belief/desire psychology literally. (Navigators talk geocentric astronomy for convenience, and nobody holds it against them; it gets them where they want to go.) The great virtue of instrumentalism—here as elsewhere—is that you get all the goodness and suffer none of the pain: you get to use propositional-attitude psychology to make behavioral predictions; you get to 'accept' all the intentional explanations that it is convenient to accept; but you don't have to answer hard questions about what the attitudes *are*.

There is, however, a standard objection to instrumentalism (again, here as elsewhere): it's hard to explain why belief/desire psychology works so well if belief/desire psychology is, as a matter of fact, not true. I propose to steer clear, throughout this essay, of general issues in the philosophy of science; in particular of issues about the status of scientific theories at large. But—as Putnam, Boyd and others have emphasized—there is surely a presumptive inference from the predictive successes of a theory to its truth; still more so when (unlike geocentric astronomy) it is the *only* predictively successful theory in the field. It's not, to put it mildly, obvious why this presumption shouldn't militate in favor of a Realist—as against an instrumentalist—construal of belief/desire explanations.

The most extensively worked-out version of instrumentalism about the attitudes in the recent literature is surely owing to D. C. Dennett. (See the papers in Dennett (1978a), especially the essay "Intentional

Systems.") Dennett confronts the 'if it isn't true, why does it work?' problem (Dennett 1981), but I find his position obscure. Here's how I *think* it goes: (a) belief/desire explanations rest on very comprehensive rationality assumptions; it's only fully rational systems that such explanations could be literally true of. These rationality assumptions are, however, generally contrary to fact; *that's* why intentional explanations can't be better than instrumental. On the other hand, (b) intentional explanations *work* because we apply them only to evolutionary successful (or other "designed") systems; and if the behavior of a system didn't at least *approximate* rationality it wouldn't *be* evolutionarily successful; what it would be is extinct.

There is a lot about this that's problematic. To begin with, it's unclear whether there really is a rationality assumption implicit in intentional explanation and whether, if there is, the rationality assumption that's required is so strong as to be certainly false. Dennett says in "Intentional Systems" (Dennett 1978c) that unless we assume rationality, we get no behavioral predictions out of belief/desire psychology since without rationality any behavior is compatible with any beliefs and desires. Clearly, however, you don't need to assume *much* rationality if all you want is *some* predictivity; perhaps you don't need to assume more rationality than organisms actually have.

Perhaps, in short, the rationality that Dennett says that natural selection guarantees is enough to support *literal* (not just instrumental) intentional ascription. At a minimum, there seems to be a clash between Dennett's principles (a) and (b) since if it *follows from* evolutionary theory that successful organisms are pretty rational, then it's hard to see how attributions of rationality to successful organisms can be construed purely instrumentally (as merely a 'stance' that we adopt towards systems whose behavior we seek to predict).

Finally, if you admit that it's a matter of fact that some agents are rational to some degree, then you have to face the hard question of how they *can* be. After all, not *everything* that's "designed" is rational even to a degree. Bricks aren't, for example; they have the wrong kind of structure. The question what sort of structure is required for rationality does, therefore, rather suggest itself and it's very unclear that that question can be answered without talking about structures of beliefs and desires; intentional psychology is the only candidate we have so far for a theory of how rationality is achieved. This suggests—what I think is true but won't argue for here—that the rational systems are a species of the intentional ones rather than the other way around. If that is so, then it is misguided to appeal to rationality in the analysis of intentionality since, in the order of explanation, the latter is the more fundamental notion. What with one thing and another, it does seem possible to doubt that a coherent instrumentalism about the attitudes is going to be forthcoming.

Second Anti-Realist Option

You could take the view that belief/desire psychology is just plain false and skip the instrumentalist trimmings. On this way of telling the Anti-Realist story, belief/desire psychology is in competition with alternative accounts of the aetiology of behavior and should be judged in the same way that the alternatives are; by its predictive successes, by the plausibility of its ontological commitments, and by its coherence with the rest of the scientific enterprise. No doubt the predictive successes of belief/desire explanations are pretty impressive—especially when they are allowed to make free use of ceteris paribus clauses. But when judged by a second and third criteria, commonsense psychology proves to be a *bad* theory; 'stagnant science' is the preferred epithet (see Paul Churchland 1981; Stich 1983). What we ought therefore to do is get rid of it and find something better.

There is, however, some disagreement as to what something better would be like. What matters here is how you feel about Functionalism. So let's have that be our next diagnostic question.

(Is everybody still with us? In case you're not, see the decision tree in figure 13.1 for the discussion so far. Auntie's motto: a place for every person; every person in his place.)

Second Question: How Do You Feel about Functionalism?

(This is a twice-told tale, so I'll be quick. For a longer review, see Fodor 1981b; Fodor, 1981c.)

It looked, in the early 1960s, as though anybody who wanted psychology to be compatible with a physicalistic ontology had a choice between some or other kind of *behaviorism* and some or other kind of *property-identity theory.* For a variety of reasons, neither of these options seemed very satisfactory (in fact, they still don't) so a small tempest brewed in the philosophical teapot.

What came of it was a new account of the type/token relation for psychological states: psychological-state tokens were to be assigned to

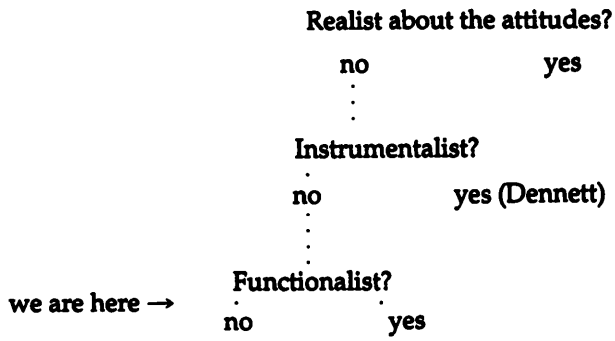

Figure 13.1 Decision Tree, stage 1.

psychological-state types *solely* by reference to their causal relations to proximal stimuli ('inputs'), to proximal responses ('outputs'), and to one another. The advertising claimed two notable virtues for this theory: first, it was *compatible* with physicalism in that it permitted tokenings of psychological states to be identical to tokenings of physical states (and thus to enjoy whatever causal properties physical states are supposed to have). Second, it permitted tokens of one and the same psychological-state type to differ arbitrarily in their physical kind. This comforted the emerging intuition that the natural domain for psychological theory might be physically heterogeneous, including a motley of people, animals, Martians (always in the philosophical literature, assumed to be silicon based), and computing machines.

Functionalism, so construed, was greeted with audible joy by the new breed of 'Cognitive Scientists' and has clearly become the received ontological doctrine in that discipline. For, if Functionalism is true, then there is plausibly a *level of explanation* between commonsense belief/desire psychology, on the one hand, and neurological (circuit-theoretic; generally 'hard-science') explanation on the other. 'Cognitive Scientists' could plausibly formulate their enterprise as the construction of theories pitched at that level. Moreover, it was possible to tell a reasonable and aesthetically gratifying story about the relations *between* the levels: commonsense belief/desire explanations *reduce* to explanations articulated in terms of functional states (at least the true ones do) because, according to Functionalism, beliefs and desires *are* functional states. And, for each (true) psychological explanation, there will be a corresponding story, to be told in hard-science terms, about how the functional states that it postulates are "realized" in the system under study. Many different hard-science stories may correspond to one and the same functional explanation since, as we saw, the criteria for the tokening of functional states abstract from the physical character of the tokens. (The most careful and convincing Functionalist manifestos I know are Block 1980; and Cummins 1983; q.v.)

Enthusiasm for Functionalism was (is) not, however, universal. For example, viewed from a neuroscientist's perspective (or from the perspective of a hard-line "type-physicalist") Functionalism may appear to be merely a rationale for making do with bad psychology. A picture many neuroscientists have is that, if there really are beliefs and desires (or memories, or percepts, or mental images or whatever else the psychologist may have in his grab bag), it ought to be possible to "find" them in the brain; where what *that* requires is that two tokens of the same *psychological* kind (today's desire to visit Auntie, say, and yesterday's) should correspond to two tokens of the same *neurological* kind (today's firing of neuron #535, say, and yesterday's). Patently, Functionalism relaxes that requirement; relaxes it, indeed, to the point of invisibility. Functionalism just *is* the doctrine that the psychologist's theoretical taxonomy doesn't need to look "natural" from the point of

view of any lower-level science. This seems to some neuroscientists, and to some of their philosopher friends, like letting psychologists get away with murder. (See, for example, Churchland 1981, which argues that Functionalism could have "saved" alchemy if only the alchemists had been devious enough to devise it.) There is, for once, something tangible at issue here: who has the right theoretical vocabulary for explaining behavior determines who should get the grants.

So much for Functionalism except to add that one can, of course, combine *accepting* the Functionalist ontology with *rejecting* the reduction of belief/desire explanations to functional ones (for example because you think that, though *some* Functionalist psychological explanations are true, no commonsense belief/desire psychological explanations are). Bearing this proviso in mind, we can put some more people in their places: if you are Anti-Realist (and anti-instrumentalist) about belief/desire psychology *and* you think there is no Functional level of explanation, then probably you think that behavioral science is (or, anyhow, ought to be) neuroscience.[1] (A fortiori, you will be no partisan of RTM, which is, of course, way over on the other side of the decision tree.) The Churchlands are the paradigm inhabitants of this niche. On the other hand, if you combine eliminativist sentiments about propositional attitudes with enthusiasm for the functional individuation of mental states, then you anticipate the eventual *replacement* of commonsense belief/desire explanations by theories couched in the vocabulary of a Functionalist psychology; replacement rather than *reduction*. You are thus led to write books with such titles as *From Folk Psychology to Cognitive Science* and are almost certainly identical to Stephen Stich.

One more word about Anti-Realism. It may strike you as odd that, whereas instrumentalists hold that belief/desire psychology works so well that we can't do anything without it, eliminativists hold that it works so badly ("stagnant science" and all that) that we can't do anything *with* it. Why, you may ask, don't these Anti-Realists get their acts together?

This is not, however, a real paradox. Instrumentalists can agree with eliminativists that *for the purposes of scientific/serious explanation* the attitudes have to be dispensed with. And eliminativists can agree with instrumentalists that for *practical* purposes, the attitudes do seem quite indispensable. In fact—and here's the point I want to stress just now—what largely motivates Anti-Realism is something deeper than the empirical speculation that belief/desire explanations won't pan out as science; it's the sense that there is something intrinsically wrong with the intentional. This is so important that I propose to leave it to the very end.

Now for the other side of the decision tree. (Presently we'll get to RTM.)

If you are a Realist about propositional attitudes, then of course you think that there are beliefs and desires. Now, on this side of the tree

too you get to decide whether to be a Functionalist or not. If you are not, then you are probably John Searle, and you drop off the edge of this paper. My own view is that RTM, construed as a species of Functionalist psychology, offers the best Realist account of the attitudes that is currently available; but this view is—to put it mildly—not universally shared. There are philosophers (many of whom like Searle, Dreyfus, and Haugeland are more or less heavily invested in Phenomenology) who are hyper-Realist about the attitudes but deeply unenthusiastic about both Functionalism and RTM. It is not unusual for such theorists to hold (a) that there *is* no currently available, satisfactory answer to the question 'how could there be things that satisfy the constraints that common sense places upon the attitudes?'; and (b) that finding an answer to this question is, in any event, not the philosopher's job. (Maybe it is the psychologist's job, or the neuroscientist's. See Dreyfus 1979; Haugeland 1978; Searle 1980.)

For how the decision tree looks now, see figure 13.2.

If you think that there are beliefs and desires, and you think that they are functional states, then you get to answer the following diagnostic question:

Third Question: Are Propositional Attitudes Monadic Functional States?

This may strike you as a *silly* question. For, you may say, since propositional attitudes are by definition relations to propositions, it follows that propositional attitudes are by definition not monadic. A propositional attitude is, to a first approximation, a *pair* of a proposition and a set of intentional systems, viz., the set of intentional systems which bear that attitude to that proposition.

That would seem to be reasonable enough. But the current (Naturalistic) consensus is that if you've gone this far you will have to go further. Something has to be said about the place of the semantic and the intentional in the natural order; it won't do to have unexplicated

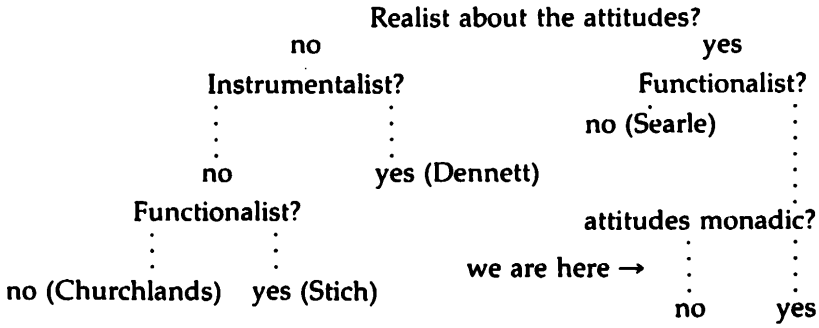

Figure 13.2 Decision Tree, stage 2.

"relations to propositions" at the foundations of the philosophy of mind.

Just *why* it won't do—precisely what physicalist or Naturalist scruples it would outrage—is, to be sure, not very clear. Presumably the issue isn't Nominalism, for why raise that issue *here;* if physicists have numbers to play with, why shouldn't psychologists have propositions? And it can't be worries about individuation since distinguishing propositions is surely no harder than distinguishing propositional attitudes and, for better or worse, we're committed to the latter on this side of the decision tree. A more plausible scruple—one I am inclined to take seriously—objects to unreduced *epistemic* relations like *grasping* propositions. One really doesn't want psychology to presuppose any of *those;* first because epistemic relations are preeminently what psychology is supposed to *explain,* and second for fear of "ontological danglers." It's not that there aren't propositions, and it's not that there aren't graspings of them; it's rather that graspings of propositions aren't plausible candidates for ultimate stuff. If they're real, they must be really something else.

Anyhow, one might as well sing the songs one knows. There *is* a reductive story to tell about *what it is* for an attitude to have a proposition as its object. So, metaphysical issues to one side, why not tell it?

The story goes as follows. Propositional attitudes are monadic, functional states of organisms. Functional states, you will recall, are type-individuated by reference to their (actual and potential) causal relations; you know everything that is essential about a functional state when you know which causal generalizations subsume it. Since, in the psychological case, the generalizations that count for type individuation are the ones that relate mental states to one another, a census of mental states would imply a network of causal interrelations. To specify such a network would be to constrain the nomologically possible mental histories of an organism; the network for a given organism would exhibit the possible patterns of causal interaction among its mental states (insofar, as least, as such patterns of interaction are relevant to the type individuation of the states). Of necessity, the actual life of the organism would appear as a path through this network.

Given the Functionalist assurance of individuation by causal role, we can assume that each mental state can be identified with a node in such a network: for each mental state there is a corresponding causal role and for each causal role there is a corresponding node. (To put the same point slightly differently, each mental state can be associated with a formula—e.g., a Ramsey sentence, see Block, 1980—that uniquely determines its location in the network by specifying its potentialities for causal interaction with each of the other mental states.) Notice, however, that while this gives a Functionalist sense to the individuation of propositional attitudes, it does not, in and of itself, say what it is for a propositional attitude to have the propositional content that it has. The

present proposal is to remedy this defect by reducing the notion of propositional content to the notion of causal role.

So far, we have a network of mental states defined by their causal interrelations. But notice that there is also a network generated by the *inferential* relations that hold among *propositions;* and it is plausible that its inferential relations are among the properties that each proposition has essentially. Thus, it is presumably a noncontingent property of the proposition that Auntie is shorter than Uncle Wilifred that it entails the proposition that Uncle Wilifred is taller than Auntie. And it is surely a noncontingent property of the proposition that *P & Q* that it entails the proposition that *P* and the proposition that *Q*. It may also be that there are evidential relations that are, in the relevant sense, noncontingent; for example, it may be constitutive of the proposition that many of the *G*'s are *F* that it is, ceteris paribus, evidence for the proposition that all of the *G*'s are *F*. If it be so, then so be it.

The basic idea is that, given the two networks—the causal and the inferential—we can establish partial isomorphisms between them. Under such an isomorphism, *the causal role of a propositional attitude mirrors the semantic role of the proposition that is its object.* So, for example, there is the proposition that John left and Mary wept; and it is partially constitutive of this proposition that it has the following semantic relations: it entails the proposition that John left; it entails the proposition that Mary wept; it is entailed by the pair of propositions {John left, Mary wept}; it entails the proposition that somebody did something; it entails the proposition that John did something; it entails the proposition that either it's raining or John left and Mary wept . . . and so forth. Likewise there are, among the potential episodes in an organism's mental life, states which we may wish to construe as: (S^1) having the belief that John left and Mary wept; (S^2) having the belief that John left; (S^3) having the belief that Mary wept; (S^4) having the belief that somebody did something; (S^5) having the belief that either it's raining or John left and Mary wept . . . and so forth. The crucial point is that it constrains the assignment of propositional contents to these mental states that the latter exhibit an appropriate pattern of causal relations. In particular, it must be true (if only under idealization) that being in S^1 tends to cause the organism to be in S^2 and S^3; that being in S^1 tends to cause the organism to be in S^4; that being (simultaneously) in states (S^2, S^3) tends—very strongly, one supposes—to cause the organism to be in state S^1, that being in state S^1 tends to cause the organism to be in state S^5 (as does being in state S^6, viz. the state of believing that it's raining). And so forth.

In short, we can make nonarbitrary assignments of propositions as the objects of propositional attitudes because there is this isomorphism between the network generated by the semantic relations among propositions and the network generated by the causal relations among men-

tal states. The assignment is nonarbitrary precisely in that it is constrained to preserve the isomorphism. And because the isomorphism is perfectly objective (which is not, however, to say that it is perfectly unique; see below), knowing what proposition gets assigned to a mental state—what the object of an attitude is—is knowing something useful. For, within the limits of the operative idealization, *you can deduce the causal consequences of being in a mental state from the semantic relations of its propositional object*. To know that John thinks that Mary wept is to know that it's highly probable that he thinks that somebody wept. To know that Sam thinks that it is raining is to know that it's highly probable that he thinks that somebody wept. To know that Sam thinks that it is raining is to know that it's highly probable that he thinks that either it is raining or that John left and Mary wept. To know that Sam thinks that it's raining and that Sam thinks that if it's raining it is well to carry an umbrella is to be far along the way to predicting a piece of Sam's behavior.

It may be, according to the present story, that preserving isomorphism between the causal and the semantic networks is *all* that there is to the assignment of contents to mental states; that nothing constrains the attribution of propositional objects to propositional attitudes *except* the requirement that isomorphism be preserved. But one need not hold that that is so. On the contrary, many—perhaps most—philosophers who like the isomorphism story are attracted by so-called 'two-factor' theories, according to which what determines the semantics of an attitude is not just its functional role but also its causal connections to objects 'in the world'. (This is, notice, still a species of functionalism since it's still causal role alone that counts for the type individuation of mental states; but two-factor theories acknowledge as semantically relevant 'external' causal relations, relations between, for example, states of the organism and *distal* stimuli. It is these mind-to-world causal relations that are supposed to determine the denotational semantics of an attitude: what it's about and what its truth-conditions are.) There are serious issues in this area, but for our purposes—we are, after all, just sightseeing—we can group the two-factor theorists with the pure functional-role semanticists.

The story I've just told you is, I think, the standard current construal of Realism about propositional attitudes.[2] I propose, therefore, to call it Standard Realism (SR for convenience). As must be apparent, SR is a compound of two doctrines: a claim about the 'internal' structure of attitudes (viz., that they are *monadic* functional states) and a claim about the source of their semantical properties (viz., that some or all of such properties arise from isomorphisms between the causal role of mental states and the implicational structure of propositions). Now, though they are usually held together, it seems clear that these claims are orthogonal. One could opt for monadic mental states without functional-role semantics; or one could opt for functional-role semantics

together with some nonmonadic account of the polyadicity of the attitudes. My own view is that SR should be rejected wholesale: that it is wrong about both the structure *and* the semantics of the attitudes. But—such is the confusion and perversity of my colleagues—this view is widely thought to be eccentric. The standard Realistic alternative to Standard Realism holds that SR is right about functional semantics but wrong about monadicity. I propose to divide these issues: monadicity first, semantics at the end.

If, in the present intellectual atmosphere, you are Realist and Functionalist about the attitudes, but you don't think that the attitudes are *monadic* functional states, then probably you think that to have a belief or a desire—or whatever—is to be related in a certain way to a Mental Representation. According to the canonical formulation of this view: for any organism *O* and for any proposition *P*, there is a relation *R* and a mental representation *MP* such that: *MP* means that (expresses the proposition that) *P;* and *O* believes that *P* iff *O* bears *R* to *MP*. (And similarly, *O* desires that *P* iff *O* bears some *different* relation, *R'*, to *MP*. And so forth. For elaboration, see Fodor 1975, 1978; Field 1978.) This is, of course, the doctrine I've been calling full-blown RTM. So we come, at last, to the bottom of the decision tree. (See figure 13.3.)

As compared with SR, RTM assumes the heavier burden of ontological commitment. It quantifies not just over such mental states as be-

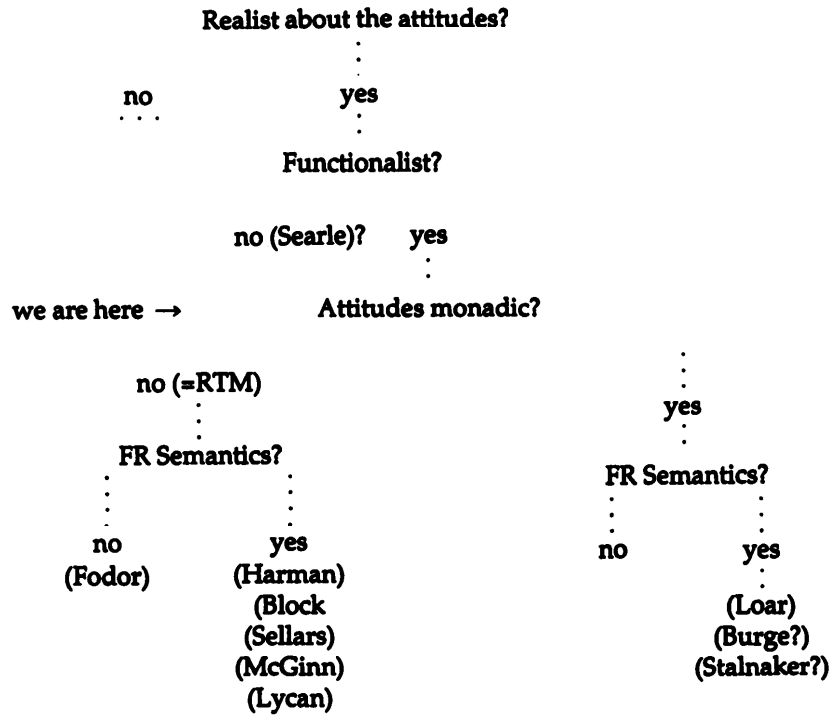

Figure 13.3 Decision Tree, stage 3.

lieving that *P* and desiring that *Q* but also over mental representations; symbols in a "language of thought." The burden of proof is thus on RTM. (Auntie holds that it doesn't matter who has the burden of proof because the choice between SR and RTM isn't a *philosophical* issue. But I don't know how she tells. Or why she cares.) There are two sorts of considerations that, in my view, argue persuasively for RTM. I think they are the implicit sources of the Cognitive Science community's commitment to the mental representation construct.

First Argument for RTM: Productivity and Constituency

The collection of states of mind is productive: for example, the thoughts that one actually entertains in the course of a mental life comprise a relatively unsystematic subset drawn from a vastly larger variety of thoughts that one could have entertained had an occasion for them arisen. For example, it has probably never occurred to you before that no grass grows on kangaroos. But, once your attention is drawn to the point, it's an idea that you are quite capable of entertaining, one which, in fact, you are probably inclined to endorse. A theory of the attitudes ought to account for this productivity; it ought to make clear what it is about beliefs and desires in virtue of which they constitute open-ended families.

Notice that Naturalism precludes saying 'there are arbitrarily many propositional attitudes because there are infinitely many propositions' and leaving it at that. The problem about productivity is that there are arbitrarily many propositional attitudes that one can *have*. Since relations between organisms and propositions aren't to be taken as primitive, one is going to have to say what it is about organic states like believing and desiring that allows them to be (roughly) as differentiated as the propositions are. If, for example, you think that attitudes are mapped to propositions in virtue of their causal roles (see above), then you have to say what it is about the attitudes that accounts for the productivity *of the set of causal roles*.

A natural suggestion is that the productivity of thoughts is like the productivity of natural languages, i.e., that there are indefinitely many thoughts to entertain for much the same reason that there are indefinitely many sentences to utter. Fine, but how do natural languages manage to be productive? Here the outlines of an answer are familiar. To a first approximation, each sentence can be identified with a certain sequence of subsentential constituents. Different sentences correspond to different ways of arranging these subsentential constituents; new sentences correspond to new ways of arranging them. And the meaning of a sentence—the proposition it expresses—is determined, in a regular way, by its constituent structure.

The constituents of sentences are, say, words and phrases. What are the constituents of propositional attitudes? A natural answer would be:

other propositional attitudes. Since, for example, you can't believe that *P* and *Q* without believing that *P* and believing that *Q*, we could take the former state to be a complex of which the latter are the relatively (or perhaps absolutely) simple parts. But a moment's consideration makes it clear that this won't work with any generality: believing that *P* or *Q* doesn't require either believing that *P* or believing that *Q*, and neither does believing that if *P* then *Q*. It looks as though we want propositional attitudes to be built out of *something*, but not out of other propositional attitudes.

There's an interesting analogy to the case of speech-acts (one of many such; see Vendler 1972). There are indefinitely many distinct assertions (i.e., there are indefinitely many propositions that one can assert); and though you can't assert that *P* and *Q* without asserting that *P* and asserting that *Q*, the disjunctive assertion, *P* or *Q*, does not imply the assertion of either of the disjuncts, and the hypothetical assertion, if *P* then *Q*, does not imply the assertion of its antecedent or its consequent. So how do you work the constituency relation for *assertions*?

Answer: you take advantage of the fact that making an assertion involves using symbols (typically it involves *uttering* symbols); the constituency relation is defined for the symbols that assertions are made by using. So, in particular, the standard (English-language) vehicle for making the assertion that either John left or Mary wept is the form of words 'either John left or Mary wept'; and, notice, this complex linguistic expression *is*, literally, a construct out of the simpler linguistic expressions 'John left' and 'Mary wept'. You can assert that *P* or *Q* without asserting that *P* or asserting that *Q*, but you can't utter the form of words '*P* or *Q*' without uttering the form of words '*P*' and the form of words '*Q*'.

The moral for treatments of the attitudes would seem to be straightforward: solve the *productivity* problem for the attitudes by appealing to constituency. Solve the *constituency* problem for the attitudes in the same way that you solve it for speech-acts: tokening an attitude involves tokening a symbol, just as tokening an assertion does. What kind of symbol do you have to token to token an attitude? A mental representation, of course. Hence RTM. (Auntie says that it is crude and preposterous and *unbiological* to suppose that people have sentences in their heads. Auntie always talks like that when she hasn't got any arguments.)

Second Argument for RTM: Mental Processes

It is possible to doubt whether, as functional-role theories of meaning would have it, the propositional contents of mental states are reducible to, or determined by, or epiphenomena of, their causal roles. But what *can't* be doubted is this: the causal roles of mental states typically closely parallel the implicational structures of their propositional objects; and

the predictive successes of propositional-attitude psychology routinely exploit the symmetries thus engendered. If we know that Psmith believes that $P \rightarrow Q$ and we know that he believes that P, then we generally expect him to infer that Q and to act according to his inference. Why do we expect this? Well, because we believe the business about Psmith to be an instance of a true and counterfactual-supporting generalization according to which believing P and believing $P \rightarrow Q$ is causally sufficient for inferring Q, ceteris paribus. But then, *what is it about the mechanisms of thinking in virtue of which such generalizations hold*? What, in particular, could believing and inferring be, such that thinking the premises of a valid inference leads, so often and so reliably, to thinking its conclusion?

It was a scandal of midcentury Anglo-American philosophy of mind that though it worried a lot about the nature of mental states (like the attitudes) it quite generally didn't worry much about the nature of mental *processes* (like thinking). This isn't, in retrospect, very surprising given the behaviorism that was widely prevalent. Mental processes are causal sequences of mental states; if you're eliminativist about the attitudes you're hardly likely to be Realist about their causal consequences. In particular, you're hardly likely to be Realist about their *causal interactions*. It now seems clear enough, however, that our theory of the structure of the attitudes *must* accommodate a theory of thinking; and that it is a preeminent constraint on the latter that it provide a mechanism for symmetry between the inferential roles of thoughts and their causal roles.

This isn't, by any means, all that easy for a theory of thinking to do. Notice, for example, that the philosophy of mind assumed in traditional British Empiricism was Realist about the attitudes and accepted a form of RTM. (Very roughly, the attitudes were construed as relations to mental images, the latter being endowed with semantic properties in virtue of what they resembled and with causal properties in virtue of their associations. Mental states were productive because complex images can be constructed out of simple ones.) But precisely because the mechanisms of mental causation were assumed to be associationistic (and the conditions for association to involve preeminently spatio-temporal propinquity), the Empiricists had no good way of connecting the *contents* of a thought with the effects of entertaining it. They therefore never got close to a plausible theory of thinking, and neither did the associationistic psychology that followed in their footsteps.

What associationism missed—to put it more exactly—was the similarity between trains of thought and *arguments*. Here, for an example, is Sherlock Holmes doing his thing at the end of "The Speckled Band":

> I instantly reconsidered my position when . . . it became clear to me that whatever danger threatened an occupant of the room could not come either from the window or the door. My attention was speedily drawn, as I have already remarked to you, to this ventilator, and to the

bell-rope which hung down to the bed. The discovery that this was a dummy, and that the bed was clamped to the floor, instantly gave rise to the suspicion that the rope was there as a bridge for something passing through the hole, and coming to the bed. The idea of a snake instantly occurred to me, and when I coupled it with my knowledge that the Doctor was furnished with a supply of the creatures from India I felt that I was probably on the right track.

The passage purports to be a bit of reconstructive psychology, a capsule history of the sequence of mental episodes which brought Holmes first to suspect, then to believe, that the Doctor did it with his pet snake. Now, back when Auntie was a girl and reasons weren't allowed to be causes, philosophers were unable to believe that such an aetiology could be literally true. I assume, however, that liberation has set in by now; we have no philosophically impressive reason to doubt that Holmes's train of thoughts went pretty much the way that he says it did.

What is therefore interesting, for our purposes, is that Holmes's story isn't *just* reconstructive psychology. It does a double duty since it also serves to assemble *premises* for a plausible inference to the *conclusion* that the doctor did it with the snake. ("A snake could have crawled through the ventilator and slithered down the bell-rope," "the Doctor was known to keep a supply of snakes in his snuff box," and so forth.) Because this train of thoughts is tantamount to an argument, Holmes expects Watson to be *convinced* by the considerations that, when they occurred to him, caused Holmes's own conviction. (Compare the sort of mental history that goes, "Well, I went to bed and slept on it, and when I woke up in the morning, I found that the problem had solved itself." Or the sort that goes, "Bell-ropes always make me think of snakes, and snakes make me think of snake oil, and snake oil makes me think of doctors; so when I saw the bell-rope it popped into my head that the Doctor and a snake might have done it between them." That's mental causation perhaps; but it's not *thinking*.)

What connects the causal-history aspect of Holmes's story with its plausible-inference aspect is precisely the parallelism between trains of thought and arguments: the thoughts that effect the fixation of the belief that P provide, often enough, good *grounds* for believing that P. (As Holmes puts it in another story, "one true inference invariably suggests others.") Were this not the case—were there not this general harmony between the semantical and the causal properties of thoughts—there wouldn't, after all, be much profit in thinking.

What you want to make thinking worth the while is that trains of thoughts should be generated by mechanisms that are generally truth-preserving (so that "a true inference [generally] suggests other inferences *that are also true*"). Argument is generally truth-preserving; that, surely, is the teleological basis of the similarity between trains of thoughts and arguments. The associationists noticed hardly any of this;

and even if they had noticed it, they wouldn't have known what to do with it. In this respect, Conan Doyle was a far deeper psychologist—far closer to what is essential about the mental life—than, say, James Joyce (or William James, for that matter).

When, therefore, Rationalist critics (including, notably, Kant) pointed out that thought—like argument—involves judging and inferring, the cat was out of the bag. Associationism was the best available form of Realism about the attitudes, and associationism failed to produce a credible mechanism for thinking. Which is to say that it failed to produce a credible theory of the attitudes. No wonder everybody gave up and turned into a behaviorist.

Cognitive Science is the art of getting the cat back in. The trick is to abandon associationism and combine RTM with the "computer metaphor." In this respect I think there really has been something like an intellectual breakthrough. Technical details to one side, this is—in my view—the *only* respect in which contemporary Cognitive Science represents a major advance over the versions of RTM that were its eighteenth- and nineteenth-century predecessors.

Computers show us how to connect semantical with causal properties *for symbols*. So, if the tokening of an attitude involves the tokening of a symbol, then we can get some leverage on connecting semantical with causal properties *for thoughts*. Here, in roughest outline, is how the story is supposed to go.

You connect the causal properties of a symbol with its semantic properties via its syntax. The syntax of a symbol is one of its second-order physical properties. To a first approximation, we can think of its syntactic structure as an abstract feature of its (geometric or acoustic) *shape*. Because, to all intents and purposes, syntax reduces to shape, and because the shape of a symbol is a potential determinant of its causal role, it is fairly easy to see how there could be environments in which the causal role of a symbol correlates with its syntax. It's easy, that is to say, to imagine symbol tokens interacting causally *in virtue of* their syntactic structures. The syntax of a symbol might determine the causes and effects of its tokenings in much the way that the geometry of a key determines which locks it will open.

But, now, we know from formal logic that certain of the semantic relations among symbols can be, as it were, "mimicked" by their syntactic relations; that, when seen from a very great distance, is what proof-theory is about. So, within certain famous limits, the semantic relation that holds between two symbols when the proposition expressed by the one is implied by the proposition expressed by the other can be mimicked by syntactic relations in virtue of which one of the symbols is derivable from the other. We can therefore build machines which have, again within famous limits, the following property: the operations of such a machine consist entirely of transformations of symbols; in the course of performing these operations, the machine is

sensitive solely to syntactic properties of the symbols; and the operations that the machine performs on the symbols are entirely confined to alterations of their shapes. Yet the machine is so devised that it will transform one symbol into another if and only if the symbols so transformed stand in certain *semantic* relations; e.g., the relation that the premises bear to the conclusion in a valid argument. Such machines—computers, of course—just *are* environments in which the causal role of a symbol token is made to parallel the inferential role of the proposition that it expresses.[3]

I expect it's clear how this is all supposed to provide an argument for quantifying over mental representations. Computers are a solution to the problem of mediating between the causal properties of symbols and their semantic properties. So *if* the mind is a sort of computer, we begin to see how you can have a theory of mental processes that succeeds where associationism (to say nothing of behaviorism) abjectly failed; a theory which explains how there could regularly be nonarbitrary content relations among causally related thoughts.

But, patently, there are going to have to be mental representations if this proposal is going to work. In computer design, causal role is brought into phase with content by exploiting parallelisms between the syntax of a symbol and its semantics. But that idea won't do the theory of *mind* any good unless there are *mental* symbols; mental particulars possessed of semantic *and syntactic* properties. There must be mental symbols because, in a nutshell, only symbols have syntax, and our best available theory of mental processes—indeed, the *only* available theory of mental processes that isn't *known* to be false—needs the picture of the mind as a syntax-driven machine.[4]

A brief addendum before we end this section: the question of the extent to which RTM must be committed to the 'explicitness' of mental representation is one that keeps getting raised in the philosophical literature (and elsewhere; see Dennett 1978b, Stabler 1983). The issue becomes clear if we consider real computers as deployed in Artificial Intelligence research. So, to borrow an example of Dennett's, there are chess machines that play as though they 'believe' that it's a good idea to get one's Queen out early. But there needn't be—in fact, there probably wouldn't be—anywhere in the system of heuristics that constitutes the program of such a machine a symbol that *means* '(try and) get your Queen out early'; rather the machine's obedience to that rule of play is, as it were, an epiphenomenon of its following many *other* rules, much more detailed, whose joint effect is that, ceteris paribus, the Queen gets out as soon as it can. The moral is supposed to be that though the contents of *some* of the attitudes it would be natural to attribute to the machine *may* be explicitly represented, none of them *have* to be, *even assuming the sort of story about how computational processes work that is supposed to motivate RTM.* So, then, what exactly *is* RTM minimally committed to by way of explicit mental representation?

The answer should be clear in light of the previous discussion. According to RTM, mental processes are transformations of mental representations. The rules which determine the course of such transformations may, but needn't, be themselves explicitly represented. But the mental contents (the 'thoughts', as it were) that get transformed *must* be explicitly presented or the theory is simply false. To put it another way: if the occurrence of a thought is an episode in a mental process, then RTM is committed to the explicit representation of the content of the thought. Or, to put it still a third way—the way they like to put it in AI—according to RTM, programs may be explicitly represented and data structures have to be.

For the sake of a simple example, let's pretend that associationism is true; we imagine that there is a principle of Association by Proximity in virtue of which thoughts of salt get associated with thoughts of pepper. The point is that even on the assumption that it subsumes mental processes, the rule 'associate by proximity' need not itself be explicitly represented; association by proximity may emerge from dynamical properties of ideas (as in Hume) or from dynamical properties of neural stuff (as in contemporary connectionism). But what *must* be explicit is the Ideas—of pepper and salt, as it might be—that get associated. For, according to the theory, mental processes are actually *causal sequences of tokenings of such Ideas;* so, no Ideas, no mental processes.

Similarly, mutatis mutandis, for the chess case. The rule 'get it out early' may be emergent out of its own implementation; out of lower-level heuristics, that is, any one of which may or may not itself be explicitly represented. But the representation of the board—of actual or possible states of play—over which such heuristics are defined *must* be explicit or the representational theory of chess playing is simply false. The theory says that a train of chess thoughts is a causal sequence of tokenings of chess representations. If, therefore, there are trains of chess thoughts but no tokenings of chess representations, it *follows* that something is not well with the theory.

So much, then, for RTM and the polyadicity of the attitudes. What about their semanticity? We proceed to our final diagnostic question:

Fourth Question: How Do You Feel about Truth-Conditions?

I remarked above that the two characteristic tenets of SR—that the attitudes are monadic and that the semanticity of the attitudes arises from isomorphisms between the causal network of mental states and the inferential network of propositions—are mutually independent. Similarly for RTM; it's not mandatory, but you are at liberty to combine RTM with functional-role (FR) semantics if you choose. Thus, you could perfectly well say: 'Believing, desiring, and so forth are relations between intentional systems and mental representations that get tokened (in their heads, as it might be). Tokening a mental representation has

causal consequences. The totality of such consequences implies a network of causal interrelations among the attitudes . . . and so on to a functional-role semantics. In any event, it's important to see that RTM needs *some* semantic story to tell if, as we have supposed, RTM is going to be Realist about the attitudes and the attitudes have their propositional objects essentially.

Which semantic story to tell is, in my view, going to be *the* issue in mental representation theory for the foreseeable future. The questions here are so difficult, and the answers so contentious, that they really fall outside the scope of this paper; I had advertised a tour of an intellectual landscape about whose topography there exists some working consensus. Still, I want to say a little about the semantic issues by way of closing. They are the piece of Cognitive Science where philosophers feel most at home; and they're where the 'philosophy of psychology' (a discipline over which Auntie is disinclined to quantify) joins the philosophy of language (which, I notice, Auntie allows me to spell without quotes).

There are a number of reasons for doubting that a functional-role semantic theory of the sort that SR proposes is tenable. This fact is currently causing something of a crisis among people who would like to be Realists about the attitudes.

In the first place—almost, by now, too obvious to mention—functional-role theories make it seem that empirical constraints must underdetermine the semantics of the attitudes. What I've got in mind here isn't the collection of worries that cluster around the 'indeterminacy of translation' thesis; if that sort of indeterminacy is to be taken seriously at all—which I doubt—then it is equally a problem for *every* Realist semantics. There are, however, certain sources of underdetermination that appear to be built into functional-role semantics as such; considerations which suggest either that there is no unique best mapping of the causal roles of mental states on to the inferential network of propositions or that, even if there is, such a mapping would nevertheless underdetermine assignments of contents to the attitudes. I'll mention two such considerations, but no doubt there are others; things are always worse than one supposes.

Idealization

The pattern of causal dispositions actually accruing to a given mental state must surely diverge very greatly from the pattern of inferences characteristic of its propositional object. We don't, for example, believe all the consequences of our beliefs; not just because we haven't got time to, and not just because everybody is at least a little irrational, but also because we surely have some false beliefs about what the consequences of our beliefs are. This amounts to saying that some substantial idealization is required if we're to get from the causal dispositions that mental states actually exhibit to the sort of causal network that we would like

to have: a causal network whose structure is closely isomorphic to the inferential network of propositions. And now the problem is to provide a noncircular justification—one which does not itself appeal to semantical or intentional considerations—for preferring *that* idealization to an infinity or so of others that ingenuity might devise. (It won't do, of course, to say that we prefer that idealization because it's the one which allows mental states to be assigned the intuitively plausible propositional objects; for the present question is precisely whether anything besides prejudice underwrites our common-sense psychological intuitions.) Probably the idealization problem arises, in some form or other, for any account of the attitudes which proposes to reduce their semantic properties to their causal ones. That, alas, is no reason to assume that the problem can be solved.

Equivalence

Functionalism guarantees that mental states are individuated by their causal roles; hence by their position in the putative causal network. But *nothing* guarantees that *propositions* are individuated by their *inferential* roles. Prima facie, it surely seems that they are not, since equivalent propositions are ipso facto identical in their inferential liaisons. Are we therefore to say that equivalent propositions are identical? Not, at least, for the psychologist's purposes, since attitudes whose propositional objects are equivalent may nevertheless differ in their causal roles. We need to distinguish, as it might be, the belief that P from the belief that *P* and (*Q* v-*Q*), hence we need to distinguish the *proposition* that P from the proposition that *P* and (*Q* v-*Q*). But surely what distinguishes these propositions is not their inferential roles, assuming that the inferential role of a proposition is something like the set of propositions it entails and is entailed by. It seems to follow that propositions are not individuated by their position in the inferential network, hence that assignments of propositional objects to mental states, if constrained only to preserve isomorphism between the networks, ipso facto underdetermine the contents of such states. There are, perhaps, ways out of such equivalence problems; 'situation semantics' (see Barwise and Perry 1983) has recently been advertising some. But all the ways out that I've heard of violate the assumptions of FR semantics; specifically, they don't identify propositions with nodes in a network of inferential roles.

In the second place, FR semantics isn't, after all, much of a panacea for Naturalistic scruples. Though it has a Naturalistic story to tell about how mental states might be paired with their propositional objects, the semantic properties of the propositions themselves are assumed, not explained. It is, for example, an intrinsic property of the proposition that Psmith is seated that it is true or false in virtue of Psmith's posture. FR semantics simply takes this sort of fact for granted. From the naturalist's point of view, therefore, it merely displaces the main worry from: 'What's the connection between an attitude and its propositional

object?' to 'What's the connection between the propositional object of an attitude and whatever state of affairs it is that makes the proposition true or false?' Or, to put much the same point slightly differently, FR semantics has a lot to say about the mind-to-proposition problem but nothing at all to say about the mind-to-world problem. In effect FR semantics is content to hold that the attitudes inherit their satisfaction-conditions from their propositional objects and that propositions have *their* satisfaction-conditions *by stipulation*.

And, in the third place, to embrace FR semantics is to raise a variety of (approximately Quinean) issues about the individuation of the attitudes; and these, as Putnam and Stich have recently emphasized, when once conjured up are not easily put down. The argument goes like this: according to FR semantic theories, each attitude has its propositional object in virtue of its position in the causal network: 'Different objects iff different loci' holds to a first approximation. Since a propositional attitude has its propositional object essentially, this makes an attitude's identity depend on the identity of its causal role. The problem is, however, that we have no criteria for the individuation of causal roles.

The usual sceptical tactic at this point is to introduce some or other form of slippery-slope argument to show—or at least to suggest—that there *couldn't be* a criterion for the individuation of causal roles that is other than arbitrary. Stich, for example, has the case of an increasingly senile woman who eventually is able to remember about President McKinley only that he was assassinated. Given that she has no *other* beliefs about McKinley—given, let's suppose, that the *only* causal consequence of her believing that McKinley was assassinated is to prompt her to produce and assent to occasional utterances of 'McKinley was assassinated' and immediate logical consequences thereof—is it clear that she in fact has *any* beliefs about McKinley at all? But if she *doesn't* have, *when, precisely, did she cease to do so*? How much causal role does the belief that McKinley was assassinated have to have to be the belief that McKinley was assassinated? And what reason is there to suppose that this question has an answer? (See Stich 1983; and also Putnam 1983.) Auntie considers slippery-slope arguments to be in dubious taste and there is much to be said for her view. Still, it looks as though FR semantics has brought us to the edge of a morass and I, for one, am not an enthusiast for wading in it.

Well then, to summarize: the syntactic theory of mental operations promises a reductive account of the *intelligence* of thought. We can now imagine—though, to be sure, only dimly and in a glass darkly—a psychology that exhibits quite complex cognitive processes as being constructed from elementary manipulations of symbols. This is what RTM, together with the computer metaphor, has brought us; and it is, in my view, no small matter. But a theory of the *intelligence* of thought does not, in and of itself, constitute a theory of thought's *intentionality*. (Compare such early papers as Dennett 1978c, where these issues are more

or less comprehensively run together, with such second thoughts as Fodor 1981a, and Cummins 1983, where they more or less aren't.) If RTM is true, the problem of the intentionality of the mental is largely—perhaps exhaustively—the problem of the semanticity of mental representations. But of the semanticity of mental representations we have, as things now stand, no adequate account.

Here ends the tour. Beyond this point there be monsters. It may be that what one descries, just there on the farthest horizon, is a glimpse of a causal/teleological theory of meaning (Stampe 1977; Dretske 1981; Fodor 1990, 1984); and it may be that the development of such a theory would provide a way out of the current mess. At best, however, it's a long way off. I mention it only to encourage such of the passengers as may be feeling queasy.

"Are you finished playing now?"
"Yes, Auntie."
"Well, don't forget to put the toys away."
"No, Auntie."

Notes

1. Unless you are an eliminativist behaviorist (say, Watson) which puts you, for present purposes, beyond the pale.

While we're at it, it rather messes up my nice taxonomy that there are philosophers who accept a Functionalist view of psychological explanation and are Realist about belief/desire psychology, but who reject the reduction of the latter to the former. In particular, they do not accept the identification of any of the entities that Functionalist psychologists posit with the propositional attitudes that common sense holds dear. (A version of this view says that functional states "realize" propositional attitudes in much the way that the physical states are supposed to realize functional ones. See, for example, Matthews 1984.)

2. This account of the attitudes seems to be in the air these days, and, as with most doctrines that are in the air, it's a little hard to be sure exactly who holds it. Far the most detailed version is in Loar 1981, though I have seen variants in unpublished papers by Tyler Burge, Robert Stalnaker, and Hartry Field.

3. Since the methods of computational psychology tend to be those of proof theory, its limitations tend to be those of formalization. Patently, this raises the well-known issue about completeness; less obviously, it connects the Cognitive Science enterprise with the Positivist program for the formalization of inductive (and, generally, nondemonstrative) styles of argument. On the second point, see Glymour 1987.

4. It is possible to combine enthusiasm for a syntactical account of mental processes with any degree of agnosticism about the attitudes—or, for that matter, about semantic evaluability itself. To claim that the mind is a "syntax-driven machine" is precisely to hold that the theory of mental processes can be set out in its entirety without reference to any of the semantical properties of mental states (see Fodor 1981a), hence without assuming that mental states *have* any semantic properties. Stephen Stich is famous for having espoused this option (Stich 1983). My way of laying out the field has put the big divide between Realism about the attitudes and its denial. This seems to me justifiable, but

admittedly it underestimates the substantial affinities between Stich and the RTM crowd. Stich's account of what a good science of behavior would look like is far closer to RTM than it is to, for example, the eliminative materialism of the Churchlands.

References

Barwise, J., and J. Perry (1983). *Situations and attitudes*. Cambridge, MA: MIT Press.

Block, N. (1980). Troubles with functionalism. In N. Block, ed., *Readings in philosophy of psychology*, vol. 1. Cambridge, MA: Harvard University Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78, 67–90.

Cummins, R. C. (1983). *The nature of psychological explanation*. Cambridge, MA: MIT Press.

Dennett, D. C. (1978a). *Brainstorms*. Cambridge, MA: Bradford Books.

Dennett, D. C. (1978b). A cure for the common code? In D. C. Dennett, *Brainstorms*. Cambridge, MA: Bradford Books.

Dennett, D. C. (1978c). Intentional systems. In D. C. Dennett, *Brainstorms*. Cambridge, MA: Bradford Books.

Dennett, D. C. (1981). True believers: The intentional stance and why it works. In A. F. Heath, ed., *Scientific explanation: Papers based on the Herbert Spencer Lectures given in the University of Oxford*. Oxford: Clarendon Press.

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Dreyfus, H. (1979) *What computers can't do*. New York: Harper & Row.

Field, H. (1978). Mental representation. *Erkenntnis* 13, 9–61.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Fodor, J. A. (1978). Propositional attitudes. *The Monist* 61, 501–523.

Fodor, J. A. (1981a). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences* 3, 63–110.

Fodor, J. A. (1981b). The mind-body problem. *Scientific American* 244, 515–531.

Fodor, J. A. (1981c). *Representations*. Cambridge, MA: MIT Press.

Fodor, J. A. (1984). Semantics, Wisconsin style. *Synthese* 59, 231–250.

Fodor, J. A. (1985). Banish discontent. In J. Butterfield, ed., *Language, mind and logic*. Cambridge: Cambridge University Press.

Fodor, J. A. (1990). Psychosemantics, or where do truth conditions come from. In W. G. Lycan, ed., *Mind and Cognition: A Reader*. Oxford: Blackwell.

Glymour, C. (1987). Android epistemology and the frame problem. In Z. Pylyshyn, ed., *The robot's dilemma: The Frame Problem in Artificial Intelligence*. Norwood, NJ: Ablex.

Haugeland, J. (1978). The nature and plausibility of cognitivism. *Behavioral and Brain Sciences* 1, 215–226.

Loar, B. (1981). *Mind and meaning*. Cambridge: Cambridge University Press.

Matthews, R. (1984). Troubles with representationalism. *Social Research* 51, 1065–1097.

Putnam, H. (1983). Computational psychology and interpretation theory. In H. Putnam, *Realism and reason*. Cambridge: Cambridge University Press.

Putnam, H. (1986). Meaning holism. In L. Hahn and P. Schilpp, eds., *The Philosophy of W. V. Quine*. La Salle, IL: Open Court.

Searle, J. R. (1980) Minds, brains, and programs. *Behavioral and Brain Sciences* 3, 417–424.

Stabler, E. (1983). How are grammars represented? *Behavioral and Brain Sciences* 6, 391–402.

Stampe, D. (1977). Towards a causal theory of linguistic representation. In P. French, T. Uehling, and H. Wettstein, eds., *Midwest Studies in Philosophy*, vol. 2. Minneapolis: University of Minnesota Press.

Stich, S. P. (1983). *From folk psychology to cognitive science*. Cambridge, MA: MIT Press.

Vendler, Z. (1983). *Res cogitans*. Ithaca: Cornell University Press.