Just What Do We Have In Mind?

LYNNE RUDDER BAKER

M any philosophers who otherwise have disparate views on the mind share a fundamental assumption. The assumption is that mental processes, or at least those that explain behavior, are wholly determined by properties of the individual whose processes they are.[1] As elaborated by Jerry Fodor, the assumption yields the widely held view called "methodological solipsism,"[2] a view developed from the plausible thought that whatever explains an individual's behavior must be "in the head" of the individual. Put this way, the assumption sounds innocent enough, perhaps even inevitable.

Nevertheless, I believe that, as it has been construed recently, the assumption is false. At the very least, it does not deserve the largely unquestioned status it enjoys, as I hope to show by a graduated series of thought experiments. I present the thought experiments as a series to expose a shared inadequacy in a variety of individualistic views, from type-type physicalism to the most sophisticated methodological solipsism; and I present them as graduated to suggest that having accepted the first relatively uncontroversial story, one has no principled place to demur later.

Although I shall discuss the thought experiments explicitly with regard to interpretations of Jerry Fodor's views, I believe that they have broad application. For example, I believe that they refute any physicalism that holds that for each distinct type of psychological state, there is a distinct type of internal, physical state; and that they refute any functionalism that holds that for each distinct type of psychological state, there is a distinct causal role.

1. METHODOLOGICAL SOLIPSISM

Fodor has developed the assumption that mental processes must be explained without reference to anything other than properties of the individual whose processes they are as a research strategy for cognitive psychology. Under the constraint of methodological solipsism, there are no psychological differences among molecularly similar individuals; relations between an organism and its environment can play no role in an explanatory psychology.[3] Fodor supports methodological solipsism by arguing that it is needed to secure intentional explanations of action—explanations in terms of an agent's beliefs, desires, intentions.

To the contrary, I shall show that methodological solipsism precludes intentional explanations, and that a solipsistic psychology, far from taking into account the way that the agent represents the world to herself, must deny that there is a fact of the matter as to whether an agent believes one thing rather than another in intuitively obvious cases.[4] The basic difficulty is that, contrary to the assumption of Fodor and other functionalists, classification of mental tokens by solipsistic computational features does not coincide with classification of mental tokens by "content," i.e., by what is expressed by 'that'-clauses of ascriptions, considered apart from truth or reference to individuals.

Mental states attributed by 'that'-clauses are individuated by the obliquely occurring expressions in the content clause. Obliquely occurring expressions are those for which substitution of coextensive expressions is impermissible. Thus, a belief that grass is green differs from a belief that snow is white in virtue of the obliquely occurring expressions in the embedded sentences. It will be useful to have a term for mental state tokens identified by obliquely occurring expressions in 'that'-clauses. Let us say that two tokens are of different *narrow semantic types* if there are semantic differences (other than truth or reference to individuals) in what is expressed by the obliquely occurring expressions in 'that'-clauses attributing them.[5]

More precisely, if 'A believes that p' and 'B believes that q' are true and 'A believes that q' is false and all the expressions in the 'that'-clauses occur obliquely, then tokens of A's believing that p and of B's believing that q are of different narrow semantic types. Because in the above 'p' abbreviates an English sentence S, to say that A believes that p is not to say that A would assent to the sentence S; A may not know English. Rather, for current purposes, a sufficient condition for A's believing that p is that there be some sentence S' in some language such that A would sincerely and comprehendingly assent to S' and S' has S as an adequate English translation. No ontological or theoretical commitment is intended; the idea of narrow semantic type that I employ can tolerate varying intuitions about, and theoretical positions on, for example, Kripke's puzzle about belief and Putnam's example of H_2O

on Earth and XYZ on Twin Earth.[6] The notion of a narrow semantic type is meant to provide an intuitive and pretheoretical means of referring to states individuated by obliquely occurring expressions in 'that'-clauses of ascriptions of those states.

Fodor has taken cognitive science to be defined by two theses:

The Content Condition. Mental states are relations to representations and mental processes are operations on representations, which are identified "opaquely" by 'that'-clauses.[7]

The Formality Condition. Mental processes are "formal" in that they apply to representations in virtue of their nonsemantic (e.g., syntactic, computational, functional, physical) properties.

Fodor's argument for the content condition is that attributions of beliefs, etc., via 'that'-clauses are required for explaining action. Fodor's main argument for the formality condition—which, he says, is tantamount to a sort of methodological solipsism (MS, 65)—is that it is "implicit in the computational model" of the mind (MS, 107), which he clearly regards as the only promising approach.[8]

Although the idea of formality must remain "intuitive and metaphoric" (MS, 64), formal properties are taken to be nonsemantic: "What makes syntactic operations a species of formal operations is that being syntactic is a way of *not* being semantic" (MS, 64). Formal operations "are the ones that are specified without reference to such semantic properties of representations as, for example, truth, reference and meaning" (MS, 64). Although representations are said to have semantic properties, it is only formal properties of representations—wholly dependent upon the "shapes" of the representations and independent of anything outside the subject—that are relevant to mental operations. Computational principles, which "apply in virtue of the form of entities in their domain,"[9] involve only formal properties. But any (contingent) properties of representations that entail anything about the external environment in which the subject is situated, or (contingent) properties that presuppose any actual relations between the subject and the environment are excluded by the formality condition as irrelevant to a scientific psychology.

The obvious examples of properties of representations that are thus irrelevant to explaining behavior are the semantic properties of truth and reference to individuals.[10] This much coincides wih pretheoretical intuitions: your child's desire to please Santa Claus may lead her to leave out cookies even though Santa Claus does not exist. Or your sister's belief that Fleet O'Foot is sure to win the Kentucky Derby may motivate her to cash in her certificate of deposit even though the belief turns out to be false. But, I shall try to show, the formality condition has much stronger consequences than such commonplace examples suggest. Indeed, it precludes the very sorts of explanations just offered. To see why, let us turn to the thought experiments.

I shall begin with an interpretation of Fodor's view that is probably too weak to be plausible, but that is suggested by his formulation at the beginning of his paper on methodological solipsism.[11] In any case, it will set the stage for more enriched interpretations. Moreover, the counterexample to this interpretation shows that many varieties of type-type physicalism are false.

2. AN ANTI-CARTESIAN MEDITATION

The first thought experiment is aimed at refuting the view that classification of mental state tokens by their actual causal relations coincides with classification of mental state tokens by narrow semantic type. Invoking non-singular, non *de re,* nonindexical attitudes, considered apart from truth, the first thought experiment will tell against the following thesis:

> (A)   If two sequences of tokens cause two tokens of a single type of bodily movement, and if the tokens in the causal sequences are, pairwise, of the same physical types, then they are, pairwise, of the same narrow semantic types.

The thought experiments draw on several plausible assumptions. First, what a sentence says depends upon what language it is in. Second, people sometimes think in words. Third, which general belief a person expresses when sincerely and comprehendingly uttering a given sentence depends upon what language the person is speaking. Fourth, just as a single physical type of ink mark may have as tokens ink marks that have different meanings in different languages, so a single physical type of audible emission may have as tokens audible emissions that have different meanings in different languages.[12]

For example, in each of two languages, there may be meaningful expressions that are phonologically and even syntactically identical (insofar as syntax is independent of semantics), but whose correct English translations differ. A sentence from one of the languages may have "Life is short" as its closest English translation, whereas a phonologically and syntactically indistinguishable sentence from the second language may have "Art is long" as its closest English translation; in that case, native speakers of the first language would typically express different beliefs when using the phonologically and syntactically identical sentences.

Questions of indeterminacy of translation do not arise; for it does not matter whether there is a unique correct translation, just that some translations would be clearly incorrect. We need only imagine a case in which there are no plausible sets of analytical hypotheses relative to which the sentence translated into English as "Life is short," for example, could be translated into English as "Art is long." Nor do questions about the nontransitivity of translation bear upon anything I have to say.

A final preliminary: Some of the individualistic views I oppose—notably, Fodor's—draw explicitly on an analogy between minds and computers, according to which mental operations are to be treated as computations on formulas; and the only constraints on interpretation of the formulas are (to put it vaguely) coherence and consistency. Without endorsing the analogy, let me here enlist its aid. One feature of computers that should lead anyone impressed with the computer analogy to accept the thought experiments is that a given computer running a given program could be interpreted differently on different days. To cite an example of Georges Rey's:

> On Wednesday [a computer] deals with the intricacies of the SALT negotiations, on Thursday it plays chess with Bobby Fischer. Now it is perfectly possible in principle for the machine to pass through type identical computational and physical states on both days. All that would normally be needed is that on both occasions the input decks be themselves type identical, and that would occur should the two problem domains be construed, as it were, isomorphically. It's just that on Wednesday the punches in the cards are interpreted (say, by Carter) to refer to Brezhnev, Vienna, and 100-megaton bombs; and on Thursday the very same punches are interpreted (say, by Spassky) to refer to moves and pieces in chess. (MS, 91)

If minds are like computers in this central respect, then there is no conceptual bar to supposing that two minds could pass through the same types of physical and computational states and yet differ in certain of their types of mental states. The thought experiments are elaborations of this idea.[13]

It will be convenient to think of the following narrative as a movie, which we join midway and which we endure for only a few minutes.

The scene is upstate New York, where a man (who, as we know from reviews, is an unhappy lover considering making the first move to patch up his broken relationship) paces up and down in front of the building where his partner lives. We hear him mutter: "Should I give the familiar signal or not? Misunderstanding calls for reconciliation. I've been misunderstood and am ready for reconciliation. Giving the old familiar signal would be just the appropriate gesture of reconciliation." Several seconds later, he rings the doorbell in the distinctive way that is the familiar signal.

The scene changes to an unidentified frontier. Now a man (who, as we know from the reviews, is a soldier considering launching a retaliatory strike) paces up and down in front of his bunker (which bears a striking resemblance to the building in New York). Because the soldier speaks an obscure dialect (not Russian or any other well-known language), what he mutters is translated into English, in accordance with standard cinematic practice: "Should I launch the attack or not? Provocation invites retaliation. I've been provoked and am prepared to retaliate. Launching an attack would be an appro-

priate gesture of retaliation." Several seconds later, he presses a button in the particular code that launches an attack.

We leave the movie here, without waiting around to see what happens. On reflection, there seems no difficulty in supposing that the scenes described could occur in "real life." Suppose that they do. Then, the soldier does not believe what the lover believes, does not want what the lover wants, does not take into consideration what the lover takes into consideration. And the intuitive psychological differences between the two cannot be reduced to differences of objects of reference "outside the head." Their psychological differences extend to their general (nonsingular, nonindexical, non-*de-re*) beliefs. Surely the beliefs that provide the materials for their episodes of practical reasoning—for example, that misunderstanding calls for reconciliation and that provocation invites retaliation—are distinct general beliefs. Indeed, the lover, with complete mastery of his language and no slips of the tongue, emphatically denies that provocation invites retaliation. The lover's deliberation concludes with a decision to give the familiar signal; the soldier's deliberation concludes with a decision to launch the attack. Therefore, construed in terms of their opaque mental states, in terms of their *de dicto* attitudes, or in terms of the content clauses of ascriptions of attitudes, the intuitive psychological states of the lover and the soldier are radically dissimilar.

Suppose, however, that the way that the soldier launches the attack is by pushing a button with his left forefinger, and the way that the lover rings the doorbell is by pushing a button with his left forefinger. In fact, suppose that the soldier and the lover each flex the same muscles in the same way to the same degree. Considered as bodily movements, what each does is a token of the same physical type. And, considered physically, the proximate causes of their muscle flexings may be supposed to be tokens of the same physical type.

Suppose further that the portion of the soldier's dialect relevant to the story is phonologically indistinguishable from English. In fact, certain well-formed expressions in his dialect are acoustically and syntactically similar to well-formed expressions in English (though, of course, they differ semantically). Considered nonsemantically, nothing about these expressions distinguishes them from English expressions. When the real-life soldier utters aloud the sentence that is correctly translated into English as "Provocation invites retaliation," his utterance sounds just like an utterance of the English sentence "Misunderstanding calls for reconciliation."[14] There may be some play in what can be a correct translation from the dialect to English, but one thing is for sure: none of the soldier's current thoughts can correctly be translated into English as thoughts concerning reconciliation or familiar signals.

Because the bodily movements of the lover and the soldier are of the

same physical types, and the proximate causes of the movements are of the same type, and their language have the odd relation just described, it is possible that, considered physically, the mental state tokens in their respective brains that constituted their episodes of practical reasoning are tokens of the same physical type.[15] For example, the physical operation that is the lover's tokening of "Should I give the familiar signal or not?" in the lover is of the same physical type as the operation that is the soldier's tokening that gets translated into English as "Should I launch the attack or not?" Of course, when the soldier scene in the movie occurs in real life, and the soldier asks his question in his dialect (not in translation), it *sounds* like a New Yorker's token of "Should I give the familiar signal or not?" The vibrations of the air waves are of the same type.

Because matters involving translation from one spoken language to another ar notoriously tricky, let me say a word more in defense of my description of the story. The key is that the soldier expresses his belief that provocation invites retaliation by uttering a sentence in his language that is syntactically and phonologically identical to the English sentence, "Misunderstanding invites reconciliation," but that is best translated into English as "Provocation invites retaliation."

We may assume that books from the soldier's dialect have been translated into English and have met all the standards of adequacy met by translations from, say, Swahili into English. In all these translations, every occurrence of what sounds like "Misunderstanding calls for reconciliation" gets translated as "Provocation invites retaliation." Since nothing is awry in the case of the soldier, and he is a native speaker of the dialect, it would be unreasonable to suppose that he uses the sentence in a nonstandard way. Moreover, there is independent evidence that "Provocation invites retaliation" is the best translation: the soldier does what one who had that belief would do in the circumstances—he launches the attack. A psychologist who knew the soldier's dialect and observed the soldier's behavior would not hesitate to attribute to him the belief that provocation invites retaliation. There is absolutely nothing fishy about such an attribution.[16]

The story depends upon no infelicity, linguistic or otherwise, on the part of either the lover or the soldier. Considered separately, there is nothing remarkable about either one or his situation. Each is a competent speaker of his language who is engaged in practical reasoning of the most straightforward sort; neither lacks any pertinent information nor makes a mistake in reasoning. There are no complications concerning tacit beliefs. There is no appeal to any intuitions about translation, other than that it is possible. There is nothing extraordinary or untoward about the situation of either party; it is only comparison of their situations in light of a certain theoretical standpoint that suggests a peculiarity.

The story of the lover and the soldier shows not only that two mental

state tokens of the same kind (e.g., belief, desire, or intention) may be of distinct narrow semantic types without differing physically, but also that two *sequences* of mental state tokens, pairwise of the same kind, may be pairwise of distinct narrow semantic types without differing physically. We can represent the practical reasoning of each as an *n*-tuple of mental state tokens. Then we have two *n*-tuples of mental state tokens

$$<m_{11}, m_{12}, \ldots m_{1n}>$$

and

$$<m_{21}, m_{22}, \ldots m_{2n}>,$$

such that, taken pairwise (e.g., m_{11} and m_{21}, m_{12} and m_{22}, etc.), the members of each pair are of the same physical type, but of different narrow semantic types.

An immediate consequence is that many varieties of type-type physicalism, even relativized to species, are false. For two subjects may be in the same type of physical state (the brain state controlling emission of certain vocables, movement of one's finger in a certain way, etc.) without being in the same type of narrow semantic state (intending to launch an attack, as opposed to intending to give a familiar signal). Insofar as one takes such states, attributed by 'that'-clauses, to be typical psychological states, then psychological states are not wholly determined by physical states of the subject.[17] Thus, the first thought experiment, which shows that sequences of tokens of a single physical type need not be sequences of tokens of a single narrow semantic type, seems to refute type-type physicalism as set out by, for example, Kim and Armstrong.[18]

3. A SECOND ANTI-CARTESIAN MEDITATION

Someone may object that although the actual states of the lover and the soldier are of the same physical types, their dispositions regarding the button-pushing behavior seem to differ; causal role is determined, in part, by counterfactuals. Suppose, the objection goes, that just before the lover gave the familiar signal, his partner appeared at the door. The lover then would not give the signal, but, say, embrace his partner instead. But given the same physical type of stimulus, the soldier would not exhibit the same type of bodily movement; he would go ahead and push the button as before. Thus, the soldier and the lover seem to differ in disposition, in which case a more sophisticated Cartesian physicalism would not fall to the thought experiment.[19]

So let us enrich the thesis under attack by considering dispositional states as well as "occurrent" states. Suppose that the lover and the soldier not only are in the same sequences of physical states causing a bodily move-

ment, but also that they have the same dispositions to make that bodily movement. The second thought experiment is aimed at showing the falsity of this thesis:

> (B)   If two sequences of tokens cause two tokens of a single type of bodily movement, and if the tokens in the causal sequences are, pairwise, of the same physical types, and if two individuals in whom those sequences occur have the same dispositions with respect to that bodily movement, then the sequences of tokens are, pairwise, of the same narrow semantic types.

My response to the objection from dispositions is to augment the story in such a way that the soldier plausibly does have the same type of physical response as the lover to the same type of physical stimulus. For example, suppose that the soldier's sister looked just like the lover's partner and that the soldier had mistakenly believed that his sister had been killed in the provoking incident. Then, if she appeared at the door, the soldier would be so glad to see her alive, that, rather than launching the attack, he would rush up to embrace her.

Of course, the story can be extended to meet objections based upon other putative differences in dispositions regarding the (physically described) behavior in question; certainly all such putative differences proposed to me have been met in this way.

Now let us see how the thought experiments so far apply to Fodor's conjunction of the content condition and the formality condition in his paper on methodological solipsism.[20] My argument will be that tokens cannot be classified by narrow semantic type without violating the formality condition. Fodor does not consider this possibility because he supposes the content condition and the formality condition to be mutually supportive. Fodor holds—and I take this to be his central theoretical claim—that:

> mental states are distinct in content only if they are relations to formally distinct mental representations; in effect, that aspects of content can be reconstructed as aspects of form, at least insofar as appeals to content figure in accounts of the mental causation of behavior. (MS, 68)[21]

Since the term 'content' has many uses, of more than one of which Fodor avails himself, let me reformulate Fodor's theoretical claim—that aspects of content can be reconstructed as aspects of form—like this:

> (T)   Two mental state tokens of the same kind (belief, desire, etc.) are of distinct psychological types only if their representations differ formally.

Recall the story of the lover and the soldier. The only differences in their representations are narrow semantic differences, which are not

mirrored by any formal differences: the lover's representation concerns reconciliation; the soldier's, retaliation. So, we have:

(T′)   Two mental state tokens of the same kind (belief, desire, etc.) may be of distinct narrow semantic types even if their representations fail to differ formally.

From (T) and (T′), it follows that:

(T″)   Two mental state tokens of the same kind (belief, desire, etc.) may be of distinct narrow semantic types, without being of distinct psychological types.

That is, the formality condition allows no distinction between the lover's belief token that misunderstanding calls for reconciliation and the soldier's belief token that provocation invites retaliation. So, if (T) is true, then such tokens are not of distinct psychological types. The lover's belief that misunderstanding calls for reconciliation is thus counted as being of the same psychological type as the soldier's belief that provocation invites retaliation (in accordance with [T]).[22] The general point suggested by the thought experiments is that, as long as these belief tokens are characterized nonsemantically (in accordance with the formality condition), then the only difference between them is that they occur in different people; differences in narrow semantic type elude anyone adhering to the formality condition.[23]

   This result alone seems decisive reason to reject the formality condition; for it seems to me that there is a gross psychological difference between believing that provocation invites retaliation and believing that misunderstanding calls for reconciliation; any theory that fails to countenance that difference is inadequate as a psychological theory. Others may take a harder line, however, and remind us that the price of a good theory is often to give up certain intuitions. So suppose that, for the sake of the theory, we take the lover and the soldier to be in the same psychological state. In that case, it is unclear that there remains any purpose in attributing beliefs (or desires, intentions) at all; such attributions would certainly be unsuitable candidates for explaining anything.

   Let me sharpen the point by imagining an alternative to the original story. Suppose that just before the episodes of practical reasoning occurred, the lover and the soldier were exchanged; so, in this alternative, unknown to either of them, their respective environments are not what they believe them to be. Since each is unaware of the switch, he reasons as before, in his own language. But in the alternative version, the lover, believing that he is giving the familiar signal, actually launches the attack. What is disconcerting is that psychological explanations can not distinguish the soldier's (deliberate) launching of the attack in the first version from the lover's (unwitting) launching of the attack in the alternative version, without violating the formality condition.

Before showing why, let me emphasize that there is no incoherence at the level of intuitive intentional explanation.[24] From the alternative version of the story, we have this information ("data"): the lover does not believe that he is launching an attack; he is in the same physical states as the soldier who (in the first version) does believe that he is launching an attack. Both, in fact, launch attacks. Intuitively, without regard for the formality condition, it is fairly clear how to give an (adequate, to my mind) intentional explanation of each of the launchings; the explanation of the lover's (unwitting) launching of the attack would have two parts—an "intentional" part, including such attributions as that he believed that he was giving the familiar signal, and a "factual" part, including such information (unavailable to the lover) as that what he took to be a doorbell was actually a triggering device.[25] So, there is nothing particularly puzzling about the case as described. Now, however, let us subject the "data" to the theory that includes the formality condition.

Suppose that Fodor explained the soldier's launching the attack as satisfying this schema:

x believes that provocation invites retaliation.

x believes that he has been provoked and desires retaliation.

x believes that the most appropriate way to retaliate is to launch an attack.

Therefore, x launches an attack.

Within the strictures of methodological solipsism, there is no way to rule out explanation of the lover's (unwitting) launching of the attack by the very same schema. Since differences in psychological state, as we are currently considering them, require formal differences, and since there are no such formal differences between the soldier's belief that he is launching an attack and the lover's belief that he is giving a familiar signal, we attribute to the lover a single psychological state regardless of whether we characterize it as a belief that he is launching an attack or as a belief that he is giving the familiar signal. But if these are attributions of a single psychological state, they must have the same place in psychological explanations of his behavior.

This result is doubly unfortunate: First, *ex hypothesi,* the lover did not believe that he was launching the attack; therefore, such a belief cannot help explain his launching the attack. And since the formality condition does not permit a distinction between attributing to the lover the belief that he was launching the attack and attributing the belief that he was giving the familiar signal, the latter belief can be no more explanatory than the former. Thus, it is doubtful that beliefs, desires, or intentions can ever be explanatory if psychological explanations conform to the formality condition.

Second, an action performed deliberately should not receive the same psychological explanation as the same type of action (under the same de-

scription, in the same external circumstances) performed unwittingly. This suggests that the psychological explanations that do conform to the formality condition are defective; for without attributions of belief, such explanations must be blind to psychological differences between doing something deliberately and doing it unwittingly. From another angle: as long as the formality condition is honored, two *incompatible* intentional explanations are equally justified.[26] Thus, it appears that the formality condition and the requirement of methodological solipsism preclude intentional explanations of action, in which case the formality condition renders practical reasoning irrelevant to what one does.

To sum up the argument against Fodor's methodological solipsism: if we take psychological states in accordance with the formality condition, we cannot coherently ascribe content to them, in which case beliefs, desires, and intentions do not count as psychological states and do not figure in psychological explanations. On the other hand, if we take psychological states in accordance with the content condition, we can attribute beliefs, desires, and intentions, but only by violating the formality condition.[27]

More generally, the stories show that even very abstract beliefs—such as that misunderstanding calls for reconciliation—fail to conform to the strictures of methodological solipsism. The identity of such beliefs depends in part upon the language-using community of the believer; so even considered apart from semantic properties of truth and reference, such beliefs are not wholly "in the head."[28]

4. A FINAL ANTI-CARTESIAN MEDITATION

There is yet a further enrichment of the functionalist thesis available to the theorist who hopes to reconcile attribution of attitudes with the formality condition. It may be postulated that whether or not an individual has a certain belief, say, is a matter of total functional organization. Since the story of the lover and the soldier may have been just a local accident, the full machine table descriptions of the two may yet reveal differences in physical-functional state. A still more stringent functionalist view would require two individuals to share all their dispositional states in order to be in the same functional state. This suggests the following thesis:

(C)   If two sequences of tokens cause two tokens of a single type of bodily movement and if the tokens in the causal sequences are, pairwise, of the same physical types, and if two individuals in whom those sequences occur have the same dispositions with respect to all their bodily movements, then the sequences are, pairwise, of the same narrow semantic types.

To refute thesis (C), I shall propose a new thought experiment, one that

will also refute the most stringent possible thesis, and with it a variety of versions of functionalist theories. The final thought experiment[29] is aimed at refuting not only (C) but also the following:

> (D)   If two sequences of tokens cause two tokens of a single type of bodily movement and if the tokens in the causal sequences are, pairwise, of the same physical types, and if they occur in two individuals who are molecule-for-molecule duplicates, then the sequences are, pairwise, of the same narrow semantic types.

I would venture to guess that most philosophers of mind today take it as a constraint on psychological theories that molecule-for-molecule duplicates must be psychological duplicates.[30] Under such a constraint, I hope to show, psychology would have nothing to do with beliefs, desires, or intentions.

Suppose that hidden in the Andes is an isolated culture—call it *Unique* —whose language and customs are very different from ours; a casual English-speaking observer would immediately be struck by the cultural differences and would not understand the language. Even so, it happens that a small fragment of the Uniques' language is acoustically indistinguishable from grammatical English. When a Unique utters what sounds like 'Jet planes are faster than trains', she is not talking about jet planes at all—nobody in Unique has the slightest idea of what a jet plane is; the Uniques are too remote even for electricity. (The Unique sentence 'Silicon chips are useful' is best rendered in English as 'Coca leaves are good to chew'.)

Now, for reasons of their own, one rather eccentric family raises their daughter in seclusion, isolated even from much of the communal life of the village. It also happens that, unlike other Uniques, the isolated daughter becomes acquainted only with that part of the Unique language that is acoustically indistinguishable from grammatical English; so, every sentence she utters in Unique has an acoustic and syntactic twin in English, which, of course, would not be a correct translation of its Unique counterpart. For example, when the isolated daughter utters what sounds like 'Berries are more abundant than bananas', she is not talking about berries and bananas, but rather about goats and chickens.

Since the Uniques are physiologically similar to us (and, of course, subject to the same physical laws), there is no difficulty in supposing that the isolated daughter has an anatomical duplicate, a molecule-for-molecule replica, in North America. The North American duplicate also has eccentric parents, who keep their daughter not only in seclusion but also on an unusual diet. Both daughters learn their languages without normal contact with the external physical and social world.[31]

At every moment throughout their short lives, the isolated daughter and her North American duplicate are in the same types of physical states.

But the Unique daughter and her North American duplicate do not share the same *de dicto* beliefs in this sense: if objects of belief are taken to be sentence-like entities in the head, the Unique daughter and her North American counterpart each have tokens of the same formal (syntactic, causal, functional, physical) type, but of distinct narrow semantic types. The argument does not require that they differ in all of their narrow semantic states, only that they differ in at least one narrow semantic state without differing in functional organization.

Consider, for example, a belief whose only physical or behavioral manifestations are verbal dispositions. (E.g., "Love means never having to say you're sorry.") Suppose that the Unique daughter complains to her mother that the rainy season, during which she has little to entertain her, is awfully long, and her mother cheers her with the thought that she should be grateful that it does not last a century, which is a much longer period of time, longer than anyone lives. In this conversation, the Unique daughter comes to believe that centuries are long, but to express her belief in Unique, she says what sounds like the English sentence, "Good ideas are rare."

At the same time, the North American daughter and her mother are having a phonologically similar conversation about the difficulty of staving off boredom when one cannot go outside. The North American mother cheers her daughter with the thought that she should be grateful that she ever has any interesting ideas about what to do, since good ideas are hard to come by. In this way, the North American comes to believe that good ideas are rare. Both the Unique daughter and the North American daughter assent when presented with the sounds, 'Good ideas are rare', but in doing so, they do not assent to the same thing. Since both beliefs—that centuries are long and that good ideas are rare—are abstract and relatively isolated from other attitudes and from nonverbal behavior, it is not difficult to suppose that with the same machine table descriptions (and hence with the same dispositions regarding their bodily movements), the girls could differ in these narrow semantic states.[32]

Although I believe that this example suffices to show that molecular duplicates may have different beliefs (considered apart from truth and reference to individuals), it should be noted that the differences in narrow semantic type need not be confined to attitudes so removed from observation of nonverbal behavior. For example, when it occurs to the North American duplicate that berries are more abundant than bananas, it occurs to the Unique daughter that goats are larger than chickens—although to assert that goats are larger than chickens in Unique, one would utter what sounds just like the English sentence, "Berries are more abundant than bananas."

Suppose that each daughter were presented with drawings of goats, chickens, berries, and bananas and were issued instructions that sound like "Pick out the berries." The North American daughter would point to the

picture of the berries. The Unique daughter would interpret the instructions to concern goats. Would she point to the picture of a goat? No. The sounds "Pick out" do not mean in Unique what they mean in English; they should be translated as "Point to the left of." So the Unique daughter, interpreting the instructions (which sound like "Pick out the berries") as "Point to the left of the goat," also points to the picture of the berries, just like her North American counterpart.

Now suppose that the drawings are presented again, but that this time the picture of the berries is to the right of the picture of the goat. On being issued the same instructions, the North American daughter again points to the berries. Again, the Unique daughter interprets the instructions as "Point to the left of the goat." Would the Unique daughter in this case fail to point to the berries—and thus diverge in her physical movements from her North American counterpart? Again, no. Small children are apt to confuse left and right, especially if they are ambidextrous as the daughters are. The Unique daughter does not yet have an adequate grasp of the concepts of left and right and so she points to the right of the goat—i.e., to the berries.

Just luck, one may reply: If the Unique daughter is presented with enough instructions, sooner or later, her bodily movements will differ from those of her North American counterpart. This consideration is not to the point, however. Since she does not yet have the same dispositions concerning spatial directions that we do, as far as her current dispositions are concerned, there are no grounds for supposing that her bodily movements will ever differ from those of the North American daughter.[33] All that is claimed is that they are molecule-for-molecule duplicates throughout their short lives; had they lived longer, they might have developed molecular differences.

One day the Unique daughter and her North American duplicate are in forests on their respective continents. Each sees an orange mushroom. Although neither has seen such a plant before, each has been told things that lead her to have certain beliefs. The Unique daughter believes that the plant has magical properties, and is curious to experience its effects; so she eats the mushroom. The North American duplicate, undergoing the same types of functional and physical states as the Unique daughter, believes that the plant is poisonous, and is curious (as children seem to be) to know what a poisonous plant tastes like; so she, too, eats the mushroom. Of course, the North American duplicate goes through the same types of physical states as the Unique daughter. Since the mushrooms are poisonous, unfortunately, the little girls both die—in exactly the same way at exactly the same time.

Because the girls had different beliefs and desires, intentional explanations of their eating the lethal mushrooms should differ. Perhaps we would trace the Unique daughter's eating the mushroom to her false belief about magical powers and the North American daughter's action to her reckless desire to know how a poisonous plant tastes. No such explanations are

allowed by the formality condition, of course. For on the formality condition, there is no psychological difference between the Unique daughter's wanting to experience magical effects and the North American duplicate's finding out how a poisonous plant tastes; nor does the formality condition permit any psychological difference between the Unique's belief that the plant has magical powers and the North American's belief that the plant is poisonous. Typed nonsemantically, the belief and desire tokens are of the same types.

On a computational psychology guided by the formality condition, the intentional explanations invoking the girls' beliefs and desires are interchangeable; but it would be clearly unacceptable to explain the Unique's eating the mushroom by reference, say, to a desire to see how a poisonous plant tastes—a desire that she did not have.

So, in terms of the physical states they instantiate and in terms of their machine table descriptions, the Unique daughter and her North American duplicate lead parallel lives; but their beliefs, desires, and intentions diverge. Therefore, molecule-for-molecule replicas can differ in their *de dicto* beliefs.

The case of the molecular replicas, I believe, refutes any view that supposes, as functionalist views typically do, that psychological states identified by narrow semantic type coincide with functional states identified by causal role. No matter how causal role is described, molecule-for-molecule duplicates have the same causal history.[34]

5. BROADER IMPLICATIONS

The upshot is that the functionalist hope of giving an account of propositional attitudes in terms of computational states is doomed.[35] Fodor remarks, "I shall simply take it for granted that you cannot save the cognitive science program by going syntactic. Either mental representations are going to honest-to-God represent, or we are going to have to find an alternative to [the representational theory of the mind]."[36] But as we have seen, given the formality condition, they cannot represent. If psychological states are construed as computational states, then beliefs, desires, intentions, and other states attributed by 'that'-clauses fail to qualify as psychological states.[37]

Let me conclude by recapitulating some of the consequences of accepting the computational view of the mind as governed by the formality condition, or for that matter, by Stich's autonomy principle.[38]

1. A computational view of the mind allows believing that *misunderstanding calls for reconciliation* and that *provocation invites retaliation* to be tokens of a single psychological type.

2. A computational view of the mind thus precludes any *intentional* explanations of action.

3. A computational view of the mind permits no explanations of any sort of actions as ordinarily described, for example, 'launching an attack', 'giving a familiar signal'. Such commonplace actions could receive no psychological explanation at all. For without recourse to beliefs, desires, and intentions, nothing remains that is capable of explaining such actions under those descriptions.
4. A computational view of the mind cannot handle the ways that action can go wrong. Without invoking beliefs and desires, there is no distinction between the lover's accidentally launching the attack and the soldier's intentionally launching the attack, nor between the Unique's eating the mushroom in the (mistaken) belief that it has magical properties, and the duplicate's eating the mushroom in the (true) belief that it is poisonous.
5. A computational view of the mind makes the whole legal process unintelligible. The determination of intentions, beliefs, and desires is integral to many legal proceedings; on the formality condition, there are no correct determinations of such attitudes.
6. A computational view of the mind allows no way to assess action as to its rationality or irrationality: Without access to what an agent believes that she is doing, there are no grounds for determining the rational merits of what she does.
7. A computational view of the mind makes moral judgments false or senseless. If there is no psychological difference between believing that one is doing A and not so believing, it is inappropriate to praise or blame a person for doing A.

All these unsavory consequences of a psychology constrained by the formality condition are aspects of what is perhaps the most fundamental consequence: A computational view of the mind severs psychological explanation from practical reasoning; since the lover's and the soldier's episodes of practical reasoning are of formally indistinguishable types, the formality condition makes a mystery of how practical reasoning can be connected to action. The deliberations of the soldier that lead up to his launching the attack cannot figure in the true psychological explanation of what he did (which could only be described in such terms as 'muscle-flexing in left forefinger'). It would thus become totally unclear how action can be connected with, for example, satisfaction of desires.

These consequences seem to me to be sufficiently devastating to make the formality condition (and with it, methodological solipsism and the computational view of the mind) utterly implausible. A computational psychology would seem useless for the aims of prediction and control as usually conceived: the only descriptions under which behavior could be predicted would be ones that hold no psychological or ethical or social or legal interest.

Such a psychology may be able to predict that a person will contract certain muscles, but not that he will write bad checks.[39]

Fodor applauds the Cartesian point that one's mental states are entirely independent of how the world actually is. But he stops too soon. For computational psychology governed by the formality condition, not only would it be true that one's mental states are independent of how the world actually is, but much more radically, it would also be true that one's mental states are independent of how the world *seems*. For how the world seems to one is a matter of one's beliefs and other attitudes attributed by 'that'-clauses. But on a theory about mental processes that conforms to the formality condition (and to methodological solipsism), such attributions become unintelligible.[40]

Notes

1. Those who share this assumption include functionalists, who take mental states to be capable of multiple physical realizations; type-type physicalists, who take types of mental states to be nothing other than types of physical states; (some) Cartesian interactionists, who take changes in mental state to cause changes in brain states; epiphenomenalists, who take changes in mental states to be caused by changes in brain states; (some) token-token physicalists, who take tokens (i.e., datable occurrences) of mental states to be identical with tokens of brain states. Although I aim to cast doubt on all these positions, there is one sort of token-token physicalism that is compatible with everything I say here: Every occurrence of a mental event may be identical with the occurrence of a physical event, as long as what counts as physical includes features of the individual's social environment. Call this position, with which I am not taking issue here, 'social supervenience', to indicate that if the mental supervenes on anything, it supervenes on the physical-cum-social, and not on states of the individual considered in isolation. Thus, methodological solipsism is appropriate as a general target of my argument.

2. The term 'methodological solipsism' in its current use originates with Putnam. See Hilary Putnam, "The Meaning of 'Meaning'," in *Mind, Language and Reality: Philosophical Papers*, vol. 2 (Cambridge, 1975, 215-71). There is an ambiguity in formulations of methodological solipsism. In Putnam's original formulation, methodological solipsism is the assumption that "no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed" (220). Subsequent formulations, according to which, for example, narrow states are what molecular duplicates share, require that narrow states must be specifiable independently of any facts about the world "outside the head." The second formulation of methodological solipsism would be violated by presupposing that there exists anything other than the individual whose mental processes are being explained; the first formulation would be violated only by presupposing the existence of a particular entity outside the head. Formulating methodological solipsism as the requirement that mental states be explained without presupposing the existence of anything outside the head is ambiguous between the two readings. Although there may be (narrow) psychological states interestingly describable without presupposing the existence of any particular entity other than the subject, my arguments will suggest that there are no (even narrower) psychological states interestingly describable without presupposing that there exists anything other than the subject. For a discussion of ambiguity of methodological solipsism from another angle, see Kent Bach, "*De Re* Belief and Methodological Solipsism," in *Thought and Object: Essays on Intentionality*, edited by Andrew Woodfield (Oxford, 1982), 121-52, esp. 123-29.

Fodor develops his position in "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," *The Behavioral and Brain Sciences* 3 (1980): 63-109. Hereafter, this article will be referred to as "MS," and subsequent citations to MS will be made in the text.

3. Stephen P. Stich's principle of autonomy is another, more precise formulation of a solipsistic restriction, according to which "any differences between organisms which do not manifest themselves as differences in their current, internal, physical states ought to be ignored by a psychological theory." *From Folk Psychology to Cognitive Science: The Case against Belief* (Cambridge, Mass., 1983), 164.

4. The consequence is that Fodor's program, in effect, collapses into Stich's. Stich adopts a "syntactic" view of the mind, according to which putative states, individuated by "content," such as believing that *p*, are not genuine psychological states. They do not figure in explanations of behavior described "autonomously," where an autonomous behavioral description is one such that "if it applies to an organism in a given setting, then it would also apply to any replica of the organism in that setting." (*From Folk Psychology to Cognitive Science,* 167.)

5. What I am calling 'narrow semantic type' is in line with at least one of Fodor's uses of 'content': In "Propositional Attitudes" (in *Representations: Philosophical Essays on the Foundations of Cognitive Science* [Cambridge, Mass., 1981], 183), Fodor takes the " 'content' of a propositional attitude, informally, to be whatever it is that the complement of the corresponding [propositional attitude]-ascribing sentence expresses." Fodor's "opaque" taxonomy of attributions of attitudes yields attitudes identified by narrow semantic type.

6. Since Putnam's Twin Earth case may be pressed as an immediate counterexample to methodological solipsism, and since Fodor seeks to rebut such a use ("Cognitive Science and the Twin-Earth Problem," *Notre Dame Journal of Formal Logic* 23 [1982], 98-118), I shall give Fodor the benefit of the doubt here; successful defense against Putnam is irrelevant to my arguments, which raise different issues.

7. 'Content condition' is my term, but it is clear that the significance of the representational theory of the mind lies in its recourse to content. To think that Marvin is melancholy, for example, is to be in a relation "to a representation the content of which is *that* Marvin is melancholy." Or again, "mental states are distinguished by the *content* of the associated representations, so we can allow for the difference between thinking that Marvin is melancholy and thinking that Sam is . . ." (MS, 63; emphases his).

8. Fodor's argument for the content condition is that "opaque" attributions of attitudes (i.e., identified by narrow semantic type) are required to explain action: "In doing our psychology, we want to attribute mental states fully opaquely because it's the fully opaque reading which tells us what the agent has in mind, and it's what the agent has in mind that causes his behavior" (MS, 67). Or again, Fodor proposes the thesis that "when we articulate the generalizations in virtue of which behavior is contingent upon mental states, it is typically an opaque construal of the mental state attributions that does the work" (MS, 66). Finally: "Nontransparent taxonomies respect the way that the organism represents the object of its propositional attitudes *to itself,* and it is this representation which functions in the causation of behavior" (MS, abstract, 63). Fodor takes transparent and opaque taxonomies of attributions of attitudes to be associated with "naturalistic" and solipsistic psychologies, respectively. The only alternative that Fodor sees to solipsistic psychology is a "naturalistic" psychology that considers subjects as embedded in their environments. Fodor regards naturalistic psychology as unfeasible because, he claims, it would have to wait upon the completion of all the other sciences to get canonical descriptions of environments, in the absence of which the project of explaining organism/environment relations is hopeless. My arguments show that opaque taxonomies of attributions of attitudes (i.e., attitudes identified by narrow semantic type) are not solipsistic.

9. Fodor, "Propositional Attitudes," 201.

10. There are other grounds, independent of the formality condition, for supposing that the truth and reference of an agent's beliefs are irrelevant to explaining behavior. See, for example, John Perry, "The Problem of the Essential Indexical," *Nous* 13 (1979): 3-22; and William G. Lycan, "Toward a Homuncular Theory of Believing," *Cognition and Brain Theory* 4, (1981): 139-59.

11. MS, 64. There is no suggestion that dispositions are required to identify a token as being of a particular semantic type. Also see Dennett ("A Cure for the Common Code?" *Brainstorms: Philosophical Essays on Mind and Psychology* [Cambridge, Mass., 1978], 104), who raises questions about Fodor's apparent commitment to "the impossible view that . . . nothing can be believed, thought about or learned without being explicitly represented." Note, however, that my arguments strike at a different point: in the case envisaged, all the beliefs are explicitly represented. Also, I have no quarrel (unlike the Churchlands, for example) with the functionalists' contention that the objects of belief are (in some sense) linguistic entities. Rather, assuming that beliefs are explicitly represented and dependent upon language, the difficulty lies in the solipsistic presuppositions about language.

12. If we add to the four assumptions a fifth—that what language one speaks is not determined solely by what's in one's head—we are close to a valid argument for the conclusion that some *de dicto* beliefs are not in the head. But rather than focus on such an abstract argument, it seems more illuminating to approach the issues more concretely through an example.

13. In "Tom Swift and His Procedural Grandmother" (*Representations: Philosophical Essays on the Foundations of Cognitive Science* [Cambridge, Mass., 1981], 204-24), Fodor considers an example of programs simulating, on the one hand, the Six-Day War and, on the other hand, a chess game: "It's a possible (though, of course, unlikely) accident that these programs should be *indistinguishable when compiled;* viz. that the [machine language] counterparts of these programs should be identical, so that the internal career of a machine running one program would be identical, step by step, to that of a machine running the other" (207; emphasis his). Also, "machines typically don't know (or care) what the programs that they run are about; all they know (or care about) is how to run their programs. This may sound cryptical or even mystical. It's not. It's merely banal" (207). Banal or not, the possibility seems to have substantial and unforeseen implications.

14. Indeed, a more contrived example would have the lover and the soldier physically overlap in such a way that a single token expresses the two beliefs. This possibility indicates that the issue is how to characterize differences between the uses of a token. David Austin made this point to me.

15. Burge's important articles make my description of the story all the more plausible. See Tyler Burge, "Individualism and the Mental," *Midwest Studies in Philosophy* 4 (1979): 73-122; "Other Bodies," in *Thought and Object: Essays on Intentionality,* edited by Andrew Woodfield (Oxford, 1982), 97-120.

16. As Arthur Danto has pointed out, there could be two languages that are phonologically similar in this way: A token of one of the languages could be acoustically indistinguishable from a token of the second, yet the best translation of the first into English is "Motherhood is sacred," and the best translation of the second is "Beans are high in protein." ("The Last Work of Art: Artworks and Real Things," reprinted from *Theoria* 39 in *Aesthetics: A Critical Anthology,* edited by George Dickie and R. J. Sclafani (New York, 1977), 551-62). The stories here are extensions of Danto's insight into the context of practical reasoning.

17. It may not be apparent that these thought experiments also refute a common brand of Cartesian interactionism. If it were supposed that the relevant mental states were "occurrent" and dispositional states of immaterial souls, rather than of brains, a similar conclusion would follow: two individuals may be in the same type of soul state without being in the same type of narrow semantic state as long as soul states are individuated without presupposing that anything exists other than the individual whose states they are. The issue of what makes the lover's soul token a token of a particular narrow semantic state is just as problematic as what makes his brain token a token of a particular narrow semantic state. In the case imagined, which narrow semantic state the lover is in is a matter of what language he speaks; but what language one speaks cannot be determined by an individual brain or soul, considered as if nothing else existed. Since invoking putative soul states does not solve the difficulty I am raising, for the

remainder of this paper, I shall not assume that there are immaterial souls. These points emerged from a conversation with Bob Hambourger.

18. The story is applicable to numerous views. See, for example, D. M. Armstrong, *A Materialist Theory of the Mind* (New York, 1968); also Jaegwon Kim, "Physicalism and the Multiple Realizability of Mental States," reprinted in *Readings in the Philosophy of Psychology*, vol. 1, edited by Ned Block (Cambridge, Mass., 1980), 234-36; David Lewis, "An Argument for the Identity Theory," reprinted in *Materialism and the Mind-Body Problem,* edited by David Rosenthal (Englewood Cliffs, N.J.; 1971), 162-71; J. J. C. Smart, *Philosophy and Scientific Realism* (New York, 1963). At times, Dennett also seems to assume that a robot that "models" a person's internal processes must have the same beliefs as the person ("A Cure for the Common Code?" 105).

19. Igal Kvart emphasized this objection to my original example.

20. It may be objected that we should only speak of functional states *within an individual,* and hence that examples considering states in different individuals are not to the point. The objection is not to the point. (1) We could just as well consider a single individual at two times. (2) The notion of functional equivalence across individuals must make sense for Fodor since he proposes methodological solipsism as research strategy for cognitive science. (3) The example concerns sequences of physically similar *tokens* that have causally similar relations; the notion of function would make no sense if it did not follow that such sequences may be functionally equivalent.

21. Fodor qualifies this somewhat by saying, "That taxonomy in respect of content *is* compatible with the formality condition, plus or minus a bit, is perhaps *the* basic idea of modern cognitive theory" (MS, 68). I do not believe that the "plus or minus a bit" affects my argument; nor do I see a better way to try to reconcile the content condition and the formality condition than the way I suggest here.

22. This result seems to undermine *any* intentional psychology that conforms to the formality condition. In particular, it refutes Fodor's conception of propositional attitudes and, with it, his view of cognitive science:

> That is, one might think of cognitive theories as filling in explanation schema of, roughly, the form: *having the attitude R to proposition P is contingently identical to being in computational relation C to the formula (or sequence of formulae) E. (The Language of Thought* [Cambridge, Mass., 1979], 77.)

If to have an attitude were to be in a certain computational state, then having the belief that provocation invites retaliation would be "contingently identical" to having the belief that misunderstanding calls for reconciliation. Although it is unclear what contingent identity comes to in this context, there is surely no possible world in which those are the same beliefs. Although this alone does not force abandonment of (T), retention of (T) (and of the formality condition) becomes quite costly. What must be given up is the view that individuation by sentence believed coincides with individuation by psychological type.

23. Dennett makes a similar point. See his "Three Kinds of Intentional Psychology," in *Reduction, Time and Reality,* edited by Richard Healey (Cambridge, 1981), 37-61, esp. 56; and "Beyond Belief," in *Thought and Object: Essays on Intentionality,* edited by Andrew Woodfield (Oxford, 1982), 1-95.

24. It is sometimes noted that explanations are pragmatic, in that what counts as an explanation of an event for one purpose may not be a suitable explanation of the same event for another purpose. Such pragmatic differences are irrelevant here. The purpose throughout will be to exhibit the connection between mental processes and behavior.

25. In *"De Re* Belief in Action" (*Philosophical Review 91,* [1982]: 363-87), I proposed such a two-stage approach to explaining intentional action; the result of the argument of the current paper is that even the first stage, in terms of the agent's point of view, is not available to the methodological solipsist. In "Why Computers Can't Act" (*American Philosophical Quarterly 18* [1981]: 157-63), I alluded to the fact that the first-person perspective is not private.

26. In addition, Fodor's account does not seem to meet his own conditions. In "Propositional Attitudes," Fodor proposes a set of conditions of adequacy on views of propositional attitudes. Of the five "a priori conditions, which, on my view, a theory of propositional attitudes ... ought to meet," (177), I believe that the theory that Fodor actually proposes in "Methodological Solipsism Considered" fails to meet at least three. The only condition clearly fulfilled by the conjunction of the content condition and the formality condition is that propositional attitudes are analyzed as relations.

27. David Austin has pointed out that my stories may be seen as an extension of inverted-spectrum objections to belief states, states that are thought most susceptible to a functionalist account. Vis-à-vis Block's examples against functionalism, David Sanford has noted that my stories are analogous to comparing China's pain to India's tickle. Cf. Shoemaker's remark that the inverted spectrum problem for functionalism is one of a class of 'qualia inversion' problems for that view. Sidney Shoemaker, "Inverted Spectrum," *Journal of Philosophy* 79 (1982.): 368, n. 10.

28. William G. Lycan has noted that a proponent of referential semantics would hold that the difference between the soldier and the lover is merely a difference of reference; if so, the solipsist may be safe because he excludes differences of reference along with truth as irrelevant to determining psychological states. I believe that I can accommodate this objection. The reason to introduce a term like 'narrow semantic type' is to have a convenient way to refer to attitudes individuated by 'that'-clauses, as they ordinarily and pretheoretically are attributed; if attitudes cannot be thus individuated without reference to individuals, I should adjust the characterization of 'narrow semantic type' accordingly. My position would not thereby be threatened. Suppose that semantics is exhausted by truth and reference; then my claim would be that what is "in the head" cannot be coherently characterized as belief that such-and-such. This is so because what is in the head does not suffice to distinguish between belief that p and belief that q, where p and q are logically nonequivalent propositions. I need not claim (nor deny) that there is a *tertium quid*—a difference in narrow semantic type that outruns sameness of physical constitution but is not merely a difference in referential semantics. But the *tertium quid* is the solipsist's best shot if he wants an intentional psychology; for without it, it is obvious that classification by 'that'-clauses violates the formality condition—a point that I am at pains to argue below. The point that I want to establish is this: no matter how narrowly one construes belief, as long as it is still recognizably belief (i.e., attributable by 'that'-clauses), it cannot be understood wholly in terms of properties of the individual whose belief it is; therefore, if behavior is to be explained only by what's in the head, it is not to be explained by belief. To put it another way, intentional psychology that aims to explain behavior cannot be solipsistic.

29. The upcoming case of the Unique daughter takes care of objections from the language of thought. It would be to no avail, for example, to claim (implausibly) that we *never* think in a natural language, and hence that the soldier's and lover's thoughts may be encoded in different Mentalese representations. I do not see how Fodor could take this route; he says: "Presumably *which* proposition an internal representation expresses—what content it has—would be complexly determined by its functional role in the organism's mental life, including, especially, the way it is connected to stimulations and responses. Functional identity of internal representations would then be criterial for their intertranslatability" ("Propositional Attitudes," 203). In any case, there are independent arguments against the possibility of a language of thought. See Patricia Churchland, "A Perspective on Mind-Brain Research," *Journal of Philosophy* 78 (1980): 185-207, especially 189, for some arguments. Also see Dennett, "A Cure for the Common Code?" 90-108, and Gilbert Harman, "Language Learning," in *Readings in the Philosophy of Psychology,* vol. 2, edited by Ned Block (Cambridge, Mass., 1981), 38-44. Moreover, to deny the possibility of molecular duplicates (as in the Unique daughter case) would be, in an important way, to give up the machine analogy.

30. Stich explicitly invokes a "replacement argument" to justify his principle of autonomy. *From Folk Psychology to Cognitive Science,* 165.

31. The supposition that one could learn a language in such circumstances is not alien to the methodological solipsist, who must assume either that one could acquire a substantial fragment of a language—enough to be said to speak a language—if one were a brain in a vat, or that learning a language lies outside the purview of psychology.

32. From the point of view of methodological solipsism, the history of the acquisition of beliefs is irrelevant. We may as well simply stipulate that each daughter has the belief attributed; evidence that each does is provided by sincere and comprehending assent to appropriate sentences. The solipsist's concern is confined to *current* internal states, the identities of which are independent of facts about their acquisition. Telling these stories is part of the general strategy throughout: to employ unexceptional, ordinary descriptions of ordinary phenomena without the strictures of solipsism, and then to see what descriptions and explanations are available to the solipsist.

33. Igal Kvart has pressed such counterfactual cases; however, he would agree, I think, that the only counterfactuals that need concern me are those grounded in the girls' actual dispositions, not in dispositions that they would have if, say, they had had wider experience or a fuller grasp of their respective languages.

34. At the Chapel Hill Philosophy Colloquium at the University of North Carolina in 1983, Fodor argued for a reinstatement of the observation/inference distinction. Certain beliefs are said to be theory-neutral in that, given similar stimulations, individuals come to have similar beliefs regardless of differences among their theoretical commitments. I do not believe that such a move would affect my arguments for several reasons. (1) It is implausible to suppose that behavior-explaining beliefs could be construed wholly in such observation terms. 'Retaliation' is a long way from observation, and it is implausible to suppose either that all those who understand the concept of retaliation "connect" it to observation in the same way, or that such general concepts fail to function in explanations. (2) Terms that could be observation terms in Fodor's sense would not be ordinary English words like 'drinkable'. At the very least, the Unique daughter case shows that there is a gap between environmental stimulus and solipsistic stimulus. (Cf. "Point to the berries.") David Sanford suggested that the observation language project begins to resemble Husserl's attempt to isolate a purely phenomenological language; what connection would such a "language" have to ones we speak?

35. Indeed, I believe that a contradiction may be formally derived from assumptions that seem central to the enterprise of using computer models to account for propositional attitudes. See my "A Farewell to Functionalism," in preparation. Patricia Kitcher remarks: "Support for the 'computer model' . . . derives largely from the belief that the software-hardware relation is a prototype for the correct model of the relation between contentful psychological states and physiological states." ("In Defense of Intentional Psychology," *Journal of Philosophy* 81 [1984]: 103.) My arguments are intended, in part, to raise questions about the machine analogy. Thus, I would side with Searle in his denial that an individual has intentional states by virture of instantiating a program, though, of course, I am less sanguine about Searle's positive view of intentionality as a purely biological phenonomenon. See, for example, John Searle, "Minds, Brains and Programs," *Behavioral and Brain Sciences* 3 (1980): 417-57.

36. Fodor, "Cognitive Science and the Twin-Earth Problem," 102.

37. If we stipulate that psychological states are whatever states cause behavior, it may be supposed that, say, 'signing one's name' may be a canonical description of behavior conforming to the formality condition, and 'intending to sign one's name' a psychological state. (Contrast 'making a contract' and 'intending to make a contract'.) What I have tried to show is that even 'signing one's name' is not a narrow enough description to conform to solipsism. The risk of proprietary uses of terms like 'behavior' is that what emerges may have no interest for anybody.

38. In my forthcoming book, I shall develop these issues more fully. Also, by setting out a view according to which the mind is, in an important sense, social, I hope to avoid relativism as well as solipsism and foundationalism. David Austin has observed that if my examples work against solipsistic views of the mental, it is not clear how adding more people would help: Why

couldn't one raise the same objections against a "social functionalism?" I shall try to address such issues in the book.

39. Of course, one may hold, with Barbara von Eckardt, for example, that cognitive psychology does not aim to provide explanations of action so much as explanations of cognitive capacities—such as the capacity to understand stories, read, reason deductively, recall common facts and so on. ("Cognitive Psychology and Principled Skepticism," *Journal of Philosophy* 81 [1984]: 67-88.) But this line is hardly satisfactory: practical reasoning is a cognitive capacity, and one that is intimately connected with intentional action. Thus, there is no begging off explanation of action by confining attention to cognitive capacities.