Notre Dame Journal of Formal Logic
Volume 23, Number 2, April 1982

Cognitive Science and the Twin-Earth Problem

J. A. FODOR

Introduction There's something odd about the history of cognitive theories. On the one hand, practically all of them, from Descartes forward, have been thoroughly committed to mental representations as explanatory constructs. But, on the other hand, a continuing critical tradition in both philosophy and psychology argues that the mental representation construct is inherently defective and cannot be made scientifically respectable. This has been going on for a long time.[1] It's a bit as though physics had developed in parallel with a line of criticism which claimed that the notion of a particle is incoherent and must be dispensed with. Surely, one would think, some sort of resolution should eventually be achieved: either the criticisms should be shown to be misdirected, or we should give up the construct criticized. One would think, too, that there ought to be some way of telling whether one's theoretical commitments are incoherent, and that three hundred years or so ought to be long enough to find out.

Anyhow, the sky is falling again. We have a cognitive science whose main tenet is that the mind is a device for the manipulation of representations. But we also have a line of philosophical criticism that goes like this: Nothing is a representation except insofar as it has representational content, and the notion *content of a mental representation* is in jeopardy. In particular, there's a new argument that is taken to show that, even if there are mental representations, and even if mental representations have contents, still the content of a mental representation is not a function of psychological variables as cognitive scientists understand such variables. So, to that extent, the notion *content of a mental representation* is not available as an explanatory construct in theories of the sort that cognitive scientists have hoped to develop.

Now, the arguments currently fluttering the dovecotes are actually rather

Received September 2, 1981; revised October 12, 1981

indirectly related to this unsettling conclusion. In fact, these arguments arose, in the first instance, out of discussions of *lexical meaning*.[2] Working out just how the issues about lexical meaning connect with the issues about mental representation is actually not easy; much of this paper will be devoted to doing so. For starters, though, here's the original argument without comment or elaboration.

Hilary Putnam (see especially [9]) imagines a place that's just like here except for certain peculiarities of microchemistry. Call this place "Twin Earth" ("Earth_2" for short). On Earth_2 they speak a language that is just like English in respect of its phonological, morphological, and syntactic properties. They call this language by a word which they pronounce /English/ and which they write "English" (but which we will write "English_2" in aid of notational perspicuousness). Since English_2 is phono-morpho-syntactically just like English, it contains a word (which we will write as "water_2") that is pronounced /water/.

The microchemical difference between Earth and Earth_2 is that, although they, like us, have a transparent fluid that they drink, sail on, wash their cars with, and refer to by the vocable /water/, and although that fluid passes all the, as it were, phenomenological tests for water (it has specific gravity 1, it freezes at zero C, and so forth), still the stuff that looks like water on Earth_2 is, in point of chemical fact, made of XYZ ($\neq H_2O$).

Putnam's intuitions about what's going on on Earth_2 run like this:

a. What English_2 speakers refer to by using the word "water_2" is not water.
b. English_2 expressions like "water_2 is wet" have different truth conditions from the homophonic expressions of English. In particular, unlike the English homophone, the truth of "water_2 is wet" does not depend upon the wetness of water. Hence,
c. "water_2" is not the same word as "water"; the two words differ in their semantic properties.

It follows trivially that English $\neq$ English_2 ("water" occurs in one language but not in the other). Since, however, we have assumed that the difference between XYZ and H_2O is the only (relevant) difference between Earth and Earth_2, it also follows that people who are as similar as you like in their physical constitution (people who are, as we shall say, *molecularly identical*[3]) may nevertheless speak different languages. Notice that it has *not* been shown to follow (yet) that molecularly identical people may mean different things by what they say; the grounds for that latter inference remain to be explored.

So much for Earth_2. Its philosophical chemistry is, I suppose, now sufficiently well publicized that we needn't develop the example in further detail. However, before we get to our main topic, which is what Putnam's case is supposed to show about the notion *content of a mental representation*, it is desirable to understand something of what cognitive scientists have wanted that notion for; what role appeals to content are supposed to play in the sorts of explanations that cognitive science seeks to provide. Hence the following section.

Why Mental Representations Have To Have Contents Quantification over—theoretical commitment to—mental representations is what cognitive science has in common with the Classical tradition in epistemology as it developed in both Rationalist and Empiricist versions. In particular, according to both Classical and current theories, behavior is typically the effect of mental processes, mental processes are typically causal sequences of mental operations, and mental operations have mental representations as their domains. This general picture, which I have elsewhere (see [3]) called the Representational Theory of Mind (RTM) is presumably well known and I shall spare the reader further exposition.

However, current cognitive science[4] differs in important respects from earlier versions of RTM. In particular, the contemporary movement is explicit in endorsing the claim that mental processes are computational: mental operations apply to mental representations in virtue of "formal" or "syntactic" (or, anyhow, nonsemantic) properties of the representations.[5] So, to the extent that one assumes that the content of a mental representation is some sort of construct out of its semantic properties, it follows that mental operations are defined without reference to the content of the representations they apply to. (I am putting this very loosely; the question of the relation between the semantic properties of a representation and its content will loom large later on and I do not want to prejudice the issue at this stage.)

Now suppose (as above) that behavior is to be explained by exhibiting its contingency upon the mental processes that cause it. And suppose that every mental process is a sequence of mental operations and that mental operations apply to mental representations in virtue of the form of the representations. Then it looks as though it does not matter whether the notion of the *content* of a mental representation is in jeopardy since it looks as though that notion is never going to be required in order to give the explanations that cognitive scientists want to give. The idea that mental operations are formal can thus be taken to imply that the content of a mental representation is a dispensible construct, at least for the purpose of cognitive science. Indeed, some philosophers have read the moral in just this way: see, for example, [12], [13], and some of the remarks in [10].

I think, however, that this line of argument is not well-advised. Roughly, the point is this: we want not just to be able to characterize the causal chains upon which behavior is contingent, but also to state such true and counterfactual supporting generalizations about the etiology of behavior as there are to be stated. But, to put the point in a nutshell, it looks as though such generalizations typically hold in virtue of intentional properties of the behaviors that they subsume, and it looks as though we shall need to advert to the content of mental representations as part of our account of the intentionality of behavior.

Notice that just about all of the familiar, rough-and-ready examples of generalizations about the way in which behavior is contingent on mental states and processes appear to make essential reference to intentional properties of the behaviors they apply to. Consider, for an instance, that paradigm of spruced-up common-sense psychology, the practical syllogism. It says something about how someone will *act* (or, if action is thwarted, what the agent will *try* to do; or, at a minimum, about the properties that the agent would

prefer his behavior to exhibit) given that he has certain specified beliefs and desires. The generalization that the practical syllogism (and, mutatis mutandis, other decision theories) articulates thus applies to behavior under intentional description; and so, in similar ways, do the rest of our folk psychology and practically all of our cognitive science.

Now, RTM proposes a two-step reduction of the intentionality of behavior to the content of mental representations. Obviously, the details are much in dispute. But, very schematically, the idea is that (step 1) for behavior to have such and such an intentional property involves its being caused by a mental state having the corresponding propositional content; and (step 2) to have a mental state with the propositional content *that P* is to be related, in a certain way, to a mental representation which expresses the proposition that *P*. *Attempts to bring it about that P* are thus explained by reference to *intentions to bring it about that P*, and intentions to bring it about that *P* are in turn explained by reference to *mental representations which*, in effect, *mean that P*. What is essential to this pattern of explanation is that the first step accounts for the intentionality of behavior by reference to the content of a (causally efficacious) propositional attitude, and the second step accounts for the content of the propositional attitude by reference to the content of a mental representation.

It is worth emphasizing that, in canonical psychological explanations of the sort that RTM contemplates, the required specifications of propositional attitudes are characteristically *de dicto* rather than *de re*. (By a *de dicto* specification of a propositional attitude, I mean approximately one in which substitution of coreferring expressions does not, in general, preserve truth unless the expressions are synonymous.) The point here is that *de re* specifications of propositional attitudes are generally too weak to support explanations of behavior when the latter is intentionally characterized. So, *de re*, Oedipus' desire to marry Jocasta = his desire to marry his mother = his desire to marry the tallest woman in Greece (assuming that Jocasta was the tallest woman in Greece at the time when Oedipus desired to marry her). But it is only the first of these specifications of what Oedipus desired (or maybe the first two, depending on how you feel about depth psychology) that figures in canonical explanations of the behavior that Oedipus tried/intended to produce.

What all this comes down to, then, is that we need the notion of the content of a mental representation to reconstruct the notion of the content of a *de dicto* propositional attitude[6]; and we need the notion of a *de dicto* propositional attitude in order to reconstruct the notion of the intentionality of behavior; and we need the notion of the intentionality of behavior in order to state a variety of psychological generalizations which appear to be (more or less) counterfactual supporting and true, and which subsume behavior in virtue of its satisfaction of intentional descriptions.

This does not, however, quite settle the issue; it still is not out of the question that some purely syntactic (formal) specification of mental representations might do the job. For example, it is conceivable that there should be some formal property (call it "*U*") that mental representations have iff they express the property of being a unicorn, and some (different) formal

property (call it "*W*") that mental representations have iff they express the property of being a witch. So, then, instead of explaining the difference between Seymor's witch hunting and his unicorn hunting by reference to differences between the *contents* of the causally implicated mental representations, we could explain it by reference to the difference between *U* and *W*. And, of course, if mental operations are indeed computational, there *is* going to have to be *some* formal difference between the mental representation(s) that mediate unicorn hunting and the one(s) that mediate witch hunting, assuming that the difference between hunting unicorns and hunting witches could show up (even counterfactually) in distinct behaviors. For, as we have seen, it is the burden of the computational theory of mental operations that only formal differences between mental representations *can* issue in distinct behaviors; mental representations that differ only in their semantic properties must ipso facto be identical in their causal roles.

There is, however, really no reason at all to suppose that there are formal doppelgangers of each feature of the contents of mental representations that we need to advert to in our accounts of the intentional properties of behavior. Positing such "type-to-type" correspondences between formal and semantic properties of mental representations involves a much stronger assumption than that each causally efficacious difference in content must correspond to *some formal difference or other*. Clearly, we have no right to build our theories of mind upon this stronger assumption since it seems entirely possible that formally quite different mental representations should be, as it were, synonymous. (Dennett has emphasized, correctly in my view, that we shall have to take this possibility especially seriously if we want ascriptions of propositional attitudes to be comparable across individuals; a fortiori if we want them to be comparable across species.) Moreover, one might feel, even if there were such coextensions between features of content and features of form, it would nevertheless be semantic facts, and not the syntactic ones, that really account for intentionality. If, for example, there is something about a mental representation that makes the behavior it causes unicorn hunting rather than witch hunting, surely it is not something about the shape of the representation; it is something that has to do with what the representation is (or purports to be) about.

So, I think that there really are only two options given the general framework of RTM. Either we assume that there are no explanatorily indispensible intentional properties of behavior (specifically, that there are no counterfactual supporting generalizations that subsume behavior in virtue of its intentional properties) or we assume that the notion of the content of a mental representation is ineliminable at least insofar as macrolevel psychological theorizing is concerned. I simply cannot take the first of these options seriously since we have—or so it seems to me—no notion of behavioral systematicity at all except the one that makes behavior systematic under intentional description. So 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 RTM.

A Puzzle about Earth$_2$ Beliefs The upshot of the discussion so far is that we need for purposes of theorizing about the intentional properties of behavior the two coordinate constructs: *de dicto specification of a propositional attitude* and *content of a mental representation*. If, then, we want to think about the implications of the Earth$_2$ examples for RTM, we need to ask *what they show about the de dicto propositional attitudes of people who talk English$_2$*. We now turn to that question.

The first thing to say is that Putnam gives us very little help here. His discussion is framed almost solely in terms of semantic and lexicographic issues: e.g., in terms of such questions as "*what does 'water$_2$' mean*?"[7] Now, one might suppose at first-blush that to settle questions about what "water$_2$" means is to settle the corresponding questions about the *de dicto* propositional attitudes of speakers of English$_2$. The reason one might suppose this is that it is very natural to assume a "Gricean" account of the relation between the meanings of linguistic forms and the *de dicto* propositional attitudes of speaker/hearers of the language that contains the forms. The idea would be that meanings are, as it were, logical constructs out of the *de dicto* propositional attitudes of speaker/hearers (e.g., out of their *de dicto* havings and recognizings of communicative intentions). There might be some uncertainty about just *which* logical construction out of propositional attitudes meanings are; but however that goes the view would be that to fix the ascription of meanings to verbal forms is to presuppose, more or less uniquely, an ascription of corresponding *de dicto* propositional attitudes to speaker/hearers.[8]

I am very much in sympathy with this sort of view, and it may be that it can be reconciled with Putnam's intuitions about what "water$_2$" means. The present point, however, is that if one does accept Putnam's intuitions, one cannot simply take the existence of a Gricean reduction of meanings to propositional attitudes for granted.[9] We will see more of this later (indeed, it will become a major theme) but here is one relevant consideration: Putnam wants to make the *extension* of a term one of the parameters of its meaning so that, presumably, "water$_2$" means XYZ together with some other stuff. And Putnam also wants to argue that speaker/hearers need not know what the terms they use refer to. It is, indeed, the conjunction of these two doctrines that Putnam expresses by the slogan "meanings aren't in the head". However, I suppose we can assume that people normally do know their own *de dicto* intentions; that *de dicto* propositional attitudes are "in the head" even if meanings are not. Surely this assumption is part and parcel of the sort of reduction of meanings to propositional attitudes that Grice proposes. But if this is right, it is hard to see how the view that "water$_2$" means (something like) XYZ could be squared with the idea that a word means what it does because speaker/hearers have the *de dicto* propositional attitudes that they have.

All this is pretty tentative. It would, for example, be possible to give up the idea that people know their own *de dicto* propositional attitudes. You might then manage a Pickwickean-Gricean reduction of meanings to communicative intentions; Pickwickean because the communicative intentions to which meanings are reduced would, in a relevant sense, not themselves be psychological states. (And, of course, you would need to find some way of

interpreting the *de re/de dicto* distinction that does not appeal to RTM; one that does not assume that *de dicto* specifications exhibit the way the content of a propositional attitude is mentally represented.)

Anyhow, for the moment I am not arguing that Putnam's views about what "water$_2$" means cannot be reconciled with Grice's views about how meanings are related to mental states. I am arguing only that you should not just assume, on Gricean grounds, that accepting Putnam's intuitions about the meanings of English$_2$ expressions closes the issue about how the communicative intentions of English$_2$ speakers ought to be described. To put the point more succinctly: so far we have arguments that molecularly identical people can speak different languages, but we have no argument for the conclusion that would make a difference to cognitive scientists; viz., that molecularly identical people can differ in their *de dicto* propositional attitudes. One way to get the latter conclusion is to assume Grice's principle that difference in *de dicto* propositional attitudes of speaker/hearers can be inferred from differences in the meanings of the linguistic forms they use. But this inference is not available to Putnam. Grice's principle cannot be assumed by a theorist who holds that meanings are not in the head; not at least without some further tinkering.

So here is the question that I am claiming Putnam's discussion leaves open and the answer to which I am claiming is essential to understanding the implications of Putnam's examples for the cognitive science project. What communicative intentions do speakers of English$_2$ use such verbal forms as "water$_2$ is wet" to express? Or, to put much the same question slightly differently: What *de dicto* belief is a speaker of English$_2$ claiming to have when he says that he believes that water$_2$ is wet? Or, to put the question slightly differently again: What statement is "water$_2$ is wet" standardly used to make in English$_2$? (Perhaps these three questions are not in fact equivalent; it hardly matters for the discussion that follows since it does seem clear that if we knew how to answer any one of them we would be well on our way to answering the rest.) I will now run through some answers that are wrong in edifying ways. In the long run, I shall be claiming that there is *no* way of answering these questions compatible with preserving Putnam's intuitions about what "water$_2$" means; hence that Putnam's intuitions must be misled.

First gambit: "*Water$_2$ is wet*" *is used to express the* de dicto *belief that water$_2$ is wet. Reply*: this proposal is unhelpful since it is part and parcel of our quandary that *we do not know which belief the belief that water$_2$ is wet is*. One way to see the difficulty is to notice that since "water$_2$" is not, strictly speaking, an expression of English, the formula "'water$_2$ is wet' expresses the belief that water$_2$ is wet" is not, strictly speaking, well formed. (Compare "'la plume est sur la table' is true iff la plume is on the table".) What we need, of course, is a convention for understanding "water$_2$" when it occurs used (as opposed to mentioned) in English sentences. The present proposal is thus unstable since what it claims about the *de dicto* propositional attitudes of English$_2$ speakers will depend entirely upon which such convention we adopt. If, for example, we decide that "water$_2$" translates as "water", then the proposal reduces to "'water$_2$ is wet' expresses the *de dicto* belief that

water is wet". If, by contrast we translate "$water_2$" as "XYZ", then the proposal reduces to "'$water_2$ is wet' expresses the *de dicto* belief that XYZ is wet"; and so forth for other possible construals of "$water_2$". These various alternatives will be discussed severally further on.

These considerations may suggest that there is some problem about construing the communicative intentions of $English_2$ speakers vis a vis "$water_2$ is wet" that does not arise when we try to explicate *our* communicative intentions with respect to the homonophonic expressions of English. And it is true, of course, that "'water is wet' expresses the belief that water is wet", though perhaps uninformative, is at least well formed. However, the appearance of asymmetry is surely spurious. When we do the *combinatorial* part of semantics, we have some justification for simply assuming that the vocabulary of the object language is available in the metalanguage of choice; we thus take the semantics of "good" and "actress" for granted and show how the semantics of "good actress" arises therefrom (see [2]). But, of course, we cannot do that when we are embarked on Putnam's project, which is precisely that of lexicographic analysis. If there *is* an issue about what "$water_2$" means, and if it is question-begging to answer that it means *$water_2$*, then surely there must be the same issue about what "water" means and it must be equally question-begging to answer that it means *water*. Maybe what this shows is just that lexicography is a mug's game; indeed, I strongly suspect that that is true. However, that line is unavailable to Putnam, whose discussion is explicitly "almost entirely about the meaning of words rather than about the meaning of sentences" ([9], p. 216). It must, in short, be obvious that if Putnam's examples make the relation between meaning and *de dicto* propositional attitudes problematic for $English_2$, they must also do so for English. The home language cannot be viewed as privileged in this sort of study.

Second gambit: "*$Water_2$ is wet*" *is used to express the* de dicto *belief that water is wet. Reply*: I take it that we can set this proposal to one side; not because it is obviously false (on the contrary, I shall eventually argue that, for all Putnam has shown, it may well be true) but rather because *if* it is true, then there is no Twin Earth problem for us to solve. As we saw above, if Putnam's example is a problem for cognitive science, that is because it seems to show that molecularly identical people can have *de dicto* propositional attitudes that differ in content. What invites this conclusion is the "Gricean" assumption that linguistic forms which differ in meaning must ipso facto differ in the propositional attitudes they are used to express. By contrast, the present proposal is that whatever may be the case with the *meanings* of "water is wet" and "$water_2$ is wet", they are used to express the *same* propositional attitude: viz., that water is wet. On this account, the only moral to be drawn from Putnam's examples would be the irrelevance of the semantics of natural language expressions to the individuation of the propositional attitudes of speaker/hearers. (This is a moral which I shall eventually endorse, though on a very narrow construal. I propose to take the sting out of it by suggesting (a) that the *content* of a linguistic expression should be distinguished from such of its semantic properties as its truth conditions;

and (b) that content *is*—though truth conditions are not—a construct out of the communicative intentions of speaker/hearers.)

Third gambit: "*Water*$_2$ *is wet*" *is used to express the* de dicto *belief that what is called "water*$_2$*" around here is wet* ("*around here*" *being used to index Earth*$_2$). *Reply*: I put this one in because some things that Putnam says about the indexicality of kind terms may suggest it. It is, however, highly implausible and I very much doubt that Putnam actually has this solution in mind. For one thing, it is too metalinguistic sounding; whatever belief "water$_2$ is wet" is used to express is surely one that animals, prelinguistic children and (nb) people who have never heard of the word "water$_2$" can share; and none of this would be so if one accepted the present proposal about what "water$_2$ is wet" is used to say. To put this point quite generally: the belief that "water$_2$ is wet" expresses must turn out, on any acceptable analysis, to be identical with the belief that water$_2$ is wet (whatever belief *that* turns out to be; see above). But, surely, it must be possible to have the belief that water$_2$ is wet without having any metalinguistic beliefs at all. The belief that water$_2$ is wet is a belief about water$_2$, not a belief about language.

Such considerations suggest that the indexicality story does not provide necessary conditions for being the belief that "water$_2$ is wet" expresses. What is more important, however, it does not provide sufficient conditions either. Consider the parallel proposal that the (English) formula "tigers have stripes" expresses the belief that what are called "tigers" around here have stripes. Well, if this is true it must follow that having the belief that what are called "tigers" around here have stripes is sufficient for having the belief that tigers have stripes; for surely, that tigers have stripes *is* the belief that "tigers have stripes" is used to express. But that's no good for the following reason: you could have the belief that what are called "tigers" around here have stripes and *not* have the belief that tigers have stripes if, for example, you happen to think that what are called "tigers" around here are pyjamas. In that case, you would have one false belief (viz., that pyjamas are called "tigers") and one true one (viz., that pyjamas have stripes), the conjunction of which is, patently, not equivalent to the belief that "tigers have stripes" is used to express.

Similarly, with bells on, for such proposals as that "water$_2$ is wet" expresses the *de dicto* belief that this stuff is wet ("this stuff" being used to index some water$_2$). You could believe that this stuff is wet while believing that this stuff is, say, tomato juice. In that case, believing that this stuff is wet would not be believing that water$_2$ is wet, even though this stuff is, as a matter of fact, both wet and water$_2$.[10]

Fourth gambit: "*Water*$_2$ *is wet*" *is used to express the* de dicto *belief that XYZ is wet. Reply*: I think that this is what a lot of philosophers would say who share Putnam's intuitions about how lexicography should be pursued. For example, Burge [1] has recently accepted the corresponding solution for a class of examples which, as he remarks, are in important respects quite similar to Putnam's but do not involve terms for natural kinds. Nevertheless, it seems clear to me that quite familiar considerations preclude taking this line. Since the reasons for denying that "water$_2$ is wet" expresses the *de dicto* belief that XYZ is wet are equally reasons for not accepting Burge's account of the

examples he investigates, it may be worth a digression to run through one of Burge's cases.

Burge asks us to accept, for purposes of argument, the following assumptions:

a. The fact that contracts need not be written (verbal contracts bind) is constitutive of our concept of contract.
b. There is an English speaker (call him Jones) whose views about contracts are much like ours except that he is misinformed about (what we would call) "verbal contracts". In particular, Jones believes that contracts must be written, hence that *soi-disant* verbal contracts are not binding.

Burge's intuition is that we ought to say in this case that Jones has the same concept of contract that we do, notwithstanding that he (Jones) denies a truth that is, by assumption, constitutive of our concept of contract. Burge also takes it (if I read him correctly) that when Jones utters (in otherwise normal circumstances) "Smith just signed a contract", his utterance should be taken to express the belief that Smith just signed a contract; i.e., to express, *inter alia*, the very concept of which *valid though not written* is, by assumption, constitutive. Burge's point is, approximately, that what concept is expressed by what you utter is determined not (or not just) by what is "in your head", but also by what concept is expressed by the homophonic utterances of other *speakers of your language*. And, of course, it is a truism that, for paradigmatic English speakers, "contract" expresses the concept *contract*.

Burge is, in my view, putting more weight on the notion *same language* than that notion will bear. As the linguists are forever reminding us, *language* and *language community*, when not merely mystical concepts, are largely geopolitical ones having much to do with who has got the gunboats. But let us grant Burge *same language* and, while we are at it, let us grant him the notion of a truth constitutive of a concept. Still, it seems to me, we cannot grant Burge his intuitions about what belief Jones uses "Smith just signed a contract" to express. For, surely, Jones expresses the same concept by "contract" when he says *that* as when he says, for example, "I deny that verbal contracts bind". But if the concept of contract expressed in this latter case is *our* concept of contract (and if, by assumption, being binding when verbal is constitutive of our concept of contract) then the belief that Jones is expressing when he denies that verbal contracts bind is explicitly self-contradictory. Specifically, the belief expressed is that what is binding when verbal is not binding when verbal. Notice, moreover, that we have to read this belief *de dicto*; it is not just that Jones believes *of* something which is as a matter of fact so and so that it is not so and so (cf. Russell's "I thought your yacht was longer than it is"). If it means anything to say that Jones has *our* concept of contract, it must mean that we should construe his utterances of "contract" in the same way we would construe our own. If, however, we do translate that way, we get self-contradictions whenever Jones says of verbal contracts what, by Burge's own assumption, Jones believes to be true of them: viz., that there aren't any. I take it, however, that there is

a principle of charity which operates to prohibit accusing one's fellows of inconsistency in this flagrant and inflammatory way.

What is common to Burge's example and Putnam's is that in each case something that is taken to be part of the meaning of an expression that speakers use is nevertheless assumed to be something that the speakers need not grasp. In Burge's example, it is a necessary truth constitutive of the meaning; in Putnam's example, it is the extension. My point is that that is all right so far, but you get into trouble with the principle of charity if you also make the assumption that, in effect, what a verbal form means is interchangeable with the concept that it expresses: in the present case, that "contract" expresses the concept *contract* or that, mutatis mutandis, "water$_2$" expresses the concept *XYZ*. For then all sorts of innocently false statements ("verbal contracts do not bind"; "water$_2$ is not XYZ") are going to be taken to render self-contradictory beliefs, and this the principle of charity forbids.

I have the impression that there are a lot of philosophers who think it is all right to say that "water$_2$" means *XYZ* or, to continue with Burge's example, that Jones uses "contract" to express the concept *contract*. It may, therefore, be worth making explicit the ingredients of the present bind and considering in some detail the various options Burge has for getting out of it.

Burge's difficulties arise from the interaction of the following five assumptions:

1. *Verbal contracts bind* is constitutive of the meaning of "contract".
2. The meaning of a word is a construct out of the concept it expresses; in particular, words that express the same concept are synonymous.
3. When Jones says "contract" he expresses the same concept that we do when we say "contract".
4. The principle of charity.
5. The intersubstitutability of synonymous expressions in *de dicto* belief (or "says that") contexts.

Of these assumptions, Burge is more or less explicitly committed to 1 and 3, and something like 2 would seem to be required if the relation *expressing the same concept as* is to be distinguished from weaker forms of semantic equivalence that words may enter into.[11] Assumptions 4 and 5 strike me as pretty plausible, but Burge might want to try ditching one or both. Neither possibility seems very attractive, however, as we are about to see.

Suppose you give up the principle that you can substitute synonyms for synonyms in *de dicto* specifications of beliefs. That would suffice to block such inferences as from "Jones believes that (says that) verbal contracts do not bind" to "Jones believes that (says that) what binds when not verbal does not bind when not verbal"; similarly, mutatis mutandis, it would serve to block the inference from "Jones$_2$ doubts that water$_2$ is XYZ" to "Jones$_2$ doubts that XYZ is XYZ". The trouble is, however, that if you do block these inferences, it is hard to see what is left of assumption 1 (or, mutatis mutandis, of the claim that "water$_2$" *means* XYZ.) Claims about lexical meaning seem to turn very largely on the issue of what substitutes for what in *de dicto* contexts.

Moreover, it will not do for Burge to just *say* that "contract" and "what is binding though verbal", although they express the same concept,[12] are

nevertheless not substitutable in *de dicto* contexts; he will have to give some account of why they are not. The trouble is, however, that the only conceivable reason why one should not be able to make such substitutions is that beliefs about contracts and beliefs about what is binding though verbal *are not identical beliefs.*[13] But one wants to ask how beliefs about contracts *could* be distinct from beliefs about what is binding though verbal if, as Burge assures us, "contract" and "what is binding though verbal" express the same concept.

Unsurprisingly, precisely the same sort of trouble arises for beliefs about $water_2$ and beliefs about XYZ. If you cannot substitute "$water_2$" for "XYZ" *salve veritate* in *de dicto* belief contexts even though, by assumption, "$water_2$" *means XYZ*, that must be because beliefs about $water_2$ and beliefs about XYZ are ipso facto different beliefs. But if we now add the (presumably untendentious) premise that "$water_2$ is wet" expresses a belief about $water_2$, we get a contradiction of the proposal we have been investigating: viz., that "$water_2$ is wet" expresses a *de dicto* belief about the wetness of XYZ. In particular, we have (1) beliefs about $water_2 \neq$ beliefs about XYZ (in order to account for the unsubstitutability of "$water_2$" for "XYZ" in *de dicto* belief contexts); (2) "$water_2$ is wet" expresses a belief about $water_2$ (by assumption); hence (3) "$water_2$ is wet" does *not* express a belief about XYZ.

Where we are is: there is a prima facie clash between the principle of charity and Burge's assumption 3, but the way out is not to jettison the substitution *salve veritate* of synonyms for synonyms in *de dicto* contexts. An obvious alternative, however, would be to give up the principle of charity in these cases. So let us look at that.

This is not a decision to be taken lightly; charity should not be confused with mere politeness. The point is—to switch the discussion back to Putnam's example—it would be unreasonable for us to take $English_2$ speakers to be expressing self-contradictions when they utter things like "$water_2$ is not XYZ". There might, for example, perfectly well be a point in the development of their chemistry when *$water_2$ is not XYZ* is the rational thing to believe given the evidence. It can, no doubt, be rational to entertain a belief that is necessarily false; but it is hard to see how one could rationally entertain a belief with the *de dicto* content *P and* $\sim P$. Could the weight of the evidence favor a contradiction?

It may be felt, however, that this sort of argument is *too* good. For, if the principle of charity precludes taking "$water_2$ is wet" to express the belief that XYZ is wet (on pain of attributing too many inconsistent *beliefs* to $English_2$ speakers) does it not also prohibit taking "$water_2$" to *mean XYZ* (on pain of attributing too many inconsistent *sayings* to $English_2$ speakers)? If "$water_2$" means something like XYZ, then it looks as though the form of words "$water_2$ is not XYZ" is going to be something like analytically false.

I am perfectly prepared to accept this argument since I am not wedded to Putnam's intuitions about the meaning of "$water_2$". If, however, you do not like it, there is a way of avoiding it. You can argue that, given Putnam's assumptions, it is not obvious that the principle of charity should be applied to *what we say* since what we say need not be, in any very direct way, an expression of what we *de dicto* believe.

If meaning is not in the head, then we are, in a certain sense, not responsible for what what we say means. In particular, we are not responsible for the consistency of what we say in the way that we are, I suppose, responsible for the consistency of our *de dicto* beliefs. Suppose that "$water_2$" means XYZ. Then, presumably, there is some sense in which the form of words "$water_2$ is not XYZ" is self-contradictory in virtue of the meanings of its constituent expressions. But it does not follow that "$water_2$ is not XYZ" expresses a self-contradictory *de dicto* belief; no doubt one contradicts *something* when one uses that form of words, but one need not be said to contradict *oneself*. That would follow only on the assumption that you can infer, Grice-wise, from the meaning of what someone says to the content of the propositional attitudes he entertains. But, as we have repeatedly had cause to remark, given Putnam's lexicographic views, that assumption cannot be taken as self-evident. On the contrary, the principle of charity forbids us to make it in this case. To put the point in a nutshell, if meaning is not in the head, then talking a language you know is a lot like talking a language you do not know; in neither case is there a direct inference from what you utter to what you believe.

I am not, of course, recommending Putnam's lexicographic intuitions; I am only saying that he has the technical option of holding onto the principle of charity for beliefs and to "$water_2$ means XYZ" by, in effect, giving up Grice's principle and refusing to permit direct inferences from what people say to what they *de dicto* believe. On the other hand, while this position is coherent it is surely unattractive since the question what it *is* that someone believes when he believes that $water_2$ is wet is still unanswered. And we are running out of candidates.

Other Options From here on, the argument will go like this. I am going to accept the intuition that Putnam's account of the meaning of "$water_2$" is primarily intended to explain: viz., that utterances of "$water_2$ is wet" on $Earth_2$ have *different truth conditions* from the homophonic utterances on Earth. But I am going to claim that this difference in truth conditions has nothing to do with the meaning of "$water_2$" (or of "water", or of "H_2O", or of "XYZ", . . . etc.). Indeed, I shall claim that it has nothing to do with *any* fact of lexicography. Here is how I propose to show this: I shall assume, contrary to the spirit of Putnam's proposals, that the belief that "$water_2$ is wet" expresses is something of the order of: the transparent, drinkable . . . stuff people sail on is wet. (I shall refer to this as the "phenomenological belief".) And I shall argue that, *even on that assumption*, you would expect tokens of "$water_2$ is wet" to have different truth conditions from tokens of the corresponding English expression. Notice that I am not claiming that "$water_2$ is wet" *does* express the phenomenological belief; I propose to remain totally agnostic on that issue. My argument will be just that the right explanation of the facts about truth conditions survives that assumption; hence that, so far as those facts are concerned, we can hold that "$water_2$ is wet" expresses the phenomenological belief if we're inclined to do so.[14]

To begin with: If "$water_2$ is wet" expresses the phenomenological belief, so too, presumably, does "water is wet". So far, then, the present analysis raises no problems for cognitive science since we no longer have

it that molecularly identical people can differ in *de dicto* propositional attitudes. But how do you capture the intuition that "$water_2$ is wet" is true in virtue of the wetness of XYZ whereas "water is wet" is true in virtue of wetness of H_2O? Or, if you do not like "true in virtue of" talk, we can put the question this way: If "water is wet" and "$water_2$ is wet" express the same *de dicto* belief, how do you account for the intuition that "water is wet iff $water_2$ is wet" is contingent?

This is, of course, where the indexical analysis did its thing; to say that "$water_2$" is used to pick out stuff of the same kind as certain (ostensively specified) stuff "around here" is to guarantee that "$water_2$ is wet" and "water is wet" are evaluated with respect to *local* samples of the wet, transparent, potable stuff that people sail on; hence utterances of the phonological form /water is wet/ get evaluated with respect to XYZ when they occur on $Earth_2$ but with respect to H_2O when they occur on Earth. We have, however, given up the indexical analysis on internal grounds and we are heuristically committed to: "water is wet" and "$water_2$ is wet" both express the phenomenological belief. Now what?

I propose to resolve the difficulty by distinguishing between the content of a belief and its truth conditions. By the content of a belief I mean approximately what we would specify if we were asked to write down its logical form, with constants for the predicate terms. So, the content of the phenomenological belief is something like: (x) (x is drinkable, transparent, sailable-on, . . . etc., only if x is wet).[15] I assume that the contents of beliefs and the contents of sentences are connected by the (Gricean) principle that sentences share the contents of the beliefs they are used to express.

The important claim is this: you cannot go directly from the content of a belief/sentence to its truth conditions (to the conditions for its evaluation); you need at least to specify the universe of discourse for the bound variables. Moreover, I want to suggest, there is a sort of Principle of Reasonableness that operates in deciding how the universe of discourse of bound variables is to be assigned, and the effect of this principle is to determine that the evaluation of universally quantified standing sentences is relevantly local. Specifically, it ensures that such sentences are evaluated in much the way they would be if they contained demonstratives. The difference between the truth conditions of "water is wet" and of "$water_2$ is wet" are thus the consequence of the application of the Principle of Reasonableness to the two cases, or so I am about to claim.

A rough formulation of the Principle of Reasonableness might go: *do not be bloody-minded in deciding what universe of discourse sentences and beliefs will be evaluated with respect to*. I shall refine this principle, slightly, a little farther on. For the moment, consider an example of its operation in respect to quantification over times. Marco Polo wrote ([7], p. 62): "Kesmur is a province distant from Bascia seven days' journey". The first point to notice is that this looks to be a universal standing sentence with bound variables ranging over times; something along the lines of: *for all pairs* (t,t'), *if* t *is the time of the start of a journey from Kesmur to Bascia and* t' *is the time of the end of that journey, then* $t' = (t + 7\ days)$. *And similarly the other way around for journeys from Bascia to Kesmur*. The second point to notice is that there

are no (explicit) indexicals, no demonstratives, none of that stuff. The third point to notice, however, is that tokens of what Marco Polo wrote are *not* evaluated with respect to literally *all* pairs of times. For example, it does not make what Marco Polo said false that we can *now* do the trip in an hour and a half by jet, or in seven seconds by manned satellite. Nor, for that matter, does it tell against the truth of what Marco Polo wrote that, in the fifth century B.C. (before they had turbocharged camels or whatever) it might have been a forty days journey; or, indeed, that the journey might then have been impossible.

What we do, in our relentless pursuit of unbloody-mindedness, is this: we evaluate what Marco Polo wrote *as though it had contained an indexical*; as though he had written something like "Kesmur is a province distant from Bascia seven days journey *by the means of transport now available*". If, in the course of making the truth conditions of tokens explicit, we now eternalize this latter sentence by replacing indexicals with names of what they index, wet get something like " 'Kesmur is a province distant from Bascia seven days' journey' is true (as tokened by Marco Polo) iff Kesmur is a province distant from Bascia seven days by the means of transport available circa 1275". Parity of analysis would, of course, apply to a tokening of the same sentence by, say, Alan Shepard; so that, for Shepard's token, the truth rule would run " 'Kesmur is a province distant from Bascia seven days' journey' is true iff Kesmur is a province distant from Bascia seven days' journey by means of transport available circa 1980". The very same standing sentence would thus be true when Marco Polo tokened it and false when Alan Shepard did. "What we have discovered", you might be inclined to say, "is that 'journey' is implicitly indexical". But, of course, we have discovered no such thing. All that has happened is that we have conscientiously avoided bloody-mindedness in deciding how implicit quantified variables shall be evaluated; in particular, we have evaluated them with respect to the universe of discourse over which the speaker may be presumed to have intended them to range. Thus does the application of the Principle of Reasonableness in determining the universe of discourse of implicit quantified variables sometimes contrive to simulate the operation of implicit indexicals, thereby misleading unwary philosophers.

Just as we evaluate implicit quantifiers with respect to the relevantly local *times*, so too we evaluate them with respect to the relevantly local *places*; and that, I claim, is what makes kind terms seem indexical. Putnam's indexical story, you will remember, went something like: "water is wet" is used to express the belief that stuff of the same kind as this (indexing water) is wet. The present view, by contrast, is that when somebody on Earth believes that water is wet, he holds a universally quantified belief with approximately the content: *all the potable, transparent sailable-on, . . . etc., kind of stuff is wet*.[16] And when somebody on Earth_2 believes that water_2 is wet, what he holds is a belief with that *very same content*. So, according to this story, the content of the belief that water is wet = the content of the belief that water_2 is wet after all. *But* establishing the identity of the contents of these beliefs is not yet establishing how their tokenings shall be evaluated (or, equivalently, what the truth conditions on the tokens are). To do the latter, we apply the Principle of Reasonableness in the form: evaluate universally quantified beliefs

about Φ-stuff with respect to relevantly local samples of Φ-stuff; specifically, evaluate universally quantified beliefs about potable, transparent, sailable-on, . . . etc., stuff with respect to relevantly local samples of stuff that is potable, transparent, sailable-on, etc. Since, as it turns out, the stuff that satisfies that description on Earth is of a *different* kind from the stuff that satisfies that description on $Earth_2$, beliefs about potable, transparent, . . . etc., stuff get evaluated in different ways in the two places. And since, to run it into the ground, the stuff that satisfies that description on Earth is H_2O and the stuff that satisfies that description on $Earth_2$ is XYZ, it turns out that Earth-wise tokens of the phenomenological belief are true iff H_2O is wet while $Earth_2$-wise tokens of that belief are true iff XYZ is wet. Which is, as the patient reader may recall, just where we wanted to get to.

I have imposed the principle that universally quantified beliefs about Φ-stuff should be evaluated with respect to "relevantly local" samples of Φ-stuff, but I have not said what relevant localness comes to. If I had to make a stab at it, I would guess that relevant localness is fundamentally an etiological notion so that what the Principle of Reasonableness is telling us to do, in this case, is to evaluate beliefs about Φ-stuff with respect to the kind of Φ-stuff that gave rise to them. Since, it turns out, the stuff that gives rise to the phenomenological belief on Earth is of a different kind from the stuff that gives rise to it on $Earth_2$, the phenomenological belief gets evaluated differently in the two places. This, however, is very tentative, and I should want to keep the issue of analyzing the notion of relevant localness clear of the issue whether we are enjoined to evaluate universal beliefs with respect to relevantly local phenomena. The latter question seems to me a good deal less murky than the former.[17]

One more word about being reasonable. It is not only informal discourse that demands circumspection in the evaluation of quantified variables. Consider the quantifiers that bind variables in lawlike statements. The fact is that we evaluate them too with respect to our "local" bits of the universe; roughly, with respect to those regions of space-time for which isotropy can reasonably be assumed.[18] It would be worse than nit-picking, it would, in fact, be bloody-minded, to object to the periodic table of elements, or to the germ theory of disease, . . . etc., on the grounds that, for all we know, they do not hold prior to the initial bang or on the other side of black holes. Nor is this exercise of reasonableness merely "optional" (to use a term that Rorty coined for a related issue). We could not say just how our nomologically bound variables should be evaluated even if we wanted to, since we do not know with what generality the laws of even our most basic sciences hold. Scientists are just like us: they get to use bound variables even though they cannot in the usual case produce a *theory* that will pick out the universe of discourse over which the variables range.

Stopping Having come all this way, a brief retrospective may make clear the structure of the argument. What there seems to be no way of doing is to preserve simultaneously:

a. Putnam's intuitions about "$water_2$"
b. the Gricean reduction of meanings to propositional attitudes

c. the principle of charity
d. the *de re/de dicto* distinction.

Roughly, to preserve the first two we must hold that /water is wet/ expresses different beliefs on Earth and $Earth_2$. But the different beliefs that commend themselves are *XYZ is wet* and *H_2O is wet*, and these beliefs can be ascribed only *de re* if the principle of charity is to be respected. To put it slightly differently, the *de re/de dicto* distinction seems to *be* the distinction between what is in the head and what is not; so we cannot both say (with Burge and Putnam) that meanings are social and say (with Grice) that meanings are logical constructs out of *de dicto* propositional attitudes.

Giving up the Gricean reduction would get us out of this, of course, but only by blunting Putnam's polemical darts. For, if meanings are not constructs out of *de dicto* propositional attitudes, it is perfectly possible that nonsynonymous formulas (such as "water is wet" and "$water_2$ is wet", assuming Putnam's lexicography) may function to express the *same de dicto* propositional attitudes. But if this is possible, then even if Putnam is right about what "water" and "$water_2$" mean, he has provided no argument that molecularly identical people can differ in what they *de dicto* believe. If, in short, he gives up Grice's principle, there would seem to be nothing in Putnam's lexicography that is relevant to the cognitive science project.

Alternatively, we could imagine giving up the distinction between *de re* and *de dicto* propositional attitudes. But that really would make the sky fall down. For one thing, believing that water is wet looks to be a different state of mind from believing that H_2O is, and we need the *de re/de dicto* distinction to say what the difference amounts to. Moreover, as I remarked above, it looks as though we are going to need the notion *content of a* de dicto *belief* if we are to have any psychology of the contingency of behavior upon mental states at all. What would it be *like* to give that up?

Anyhow, it is just as well that we do not have to. What we can do instead is account for the fact that /water is wet/ has different truth conditions here and on $Earth_2$ by appealing to the operation of what are essentially pragmatic rather than semantic considerations: viz., by appealing to the Principle of Reasonableness rather than the putative nonsynonymy of "water" and "$water_2$". This leaves the lexicographic issues wide open; and if it turns out that such issues have no principled resolutions, maybe that is all right too.

Here are some questions that people have asked me about this paper:

1. "Aren't you saying that 'water' isn't a rigid designator for H_2O?" Answer: no, I am not saying that. What I do claim is that, if the evidence that "water" rigidly designates H_2O is the difference in truth conditions between "water is wet" and "$water_2$ is wet", then we have *no* evidence for the rigid designation claim. For, as we have seen, that difference in truth conditions is compatible with "water" and "$water_2$" both expressing the phenomenological concept.

On the other hand, the fact that "water is not H_2O" is not of the *de dicto* form *P and ~P* is *consistent* with "water" being a rigid designator for H_2O. What would make them inconsistent is the principle that if two terms rigidly designate the same thing then they have the same meaning. But we have

independent reason for doubting that this is so; cf. "equilateral triangle" and "equiangular triangle", both of which designate equilateral triangles in all possible worlds.

A fortiori, I have no argument (and no grudge) against the claim that "water is H_2O" is metaphysically necessary, since I take it that the claim is to be defended, if at all, on grounds independent of lexicography or semantics.

2. "Yes, but: is meaning in the head?" Answer: yes and no. What determines *behavior* (things like what I have called belief contents) is in the head, but you cannot get truth conditions directly out of what determines behavior. You need pragmatic principles to tell you such things as how to evaluate bound variables. On the other hand:

a. Barring explicit indexicals,[19] none of what is outside the head but relevant to the determination of truth conditions has been shown to be *specific to particular lexical items*; so it is unclear that $Earth_2$ examples have any implications for lexicography, or for what you have to learn to learn a word, or for what you have to know to know a word, . . . etc.
b. Though what is in the head does not determine extension *by itself*, it does determine extension in the context of appropriate, general pragmatic principles. So, contrary to what Putnam often suggests, there is no reason why the usual semantics of truth and reference should not apply when we map from belief contents onto the world; and, if there ever was any reason to trust evidence about extensions in establishing inferences about intensions, these reasons survive Putnam's examples. Only, when you make such inferences, keep the pragmatic effects in mind.
c. We have not had to tell a story about "linguistic division of labor". No doubt communication is a cooperative enterprise and if we insist upon being unreasonable with one another, the thing is not going to work. But nothing about the way that "water" and "$water_2$" operate suggests that their meanings have somehow been delivered into the hands of the experts. From what I have seen of how experts use language, I would be disinclined to trust them with it.
d. Individualism is all right if the notion of *de dicto* beliefs is all right; and the notion of *de dicto* belief is all right if and only if we can indeed explain the behavior of organisms by reference to the contents of their propositional attitudes. That has been clear for a long time and despite the present excitements, nothing much would seem to have changed around here.

You can come back, Chicken Little; I expect that everything is going to be all right.

NOTES

1. Compare, for example, [11] with [5].

2. Though several recent papers have sought to make the putative morals for cognitive science explicit; see [12], [13], [8], and [1].

3. Of course they are not *really*, since, as Geach [4] points out, there are molecules of H_2O in us but not in them. I take it that this detail will be seen not to prejudice the spirit of Putnam's example.

4. From here on, "cognitive science" usually denotes not a problem domain but a body of doctrine; specifically, a theory whose main tenets are RTM together with the claim that mental processes are computational. The locution "cognitive science" should thus be construed on the model of, say, "Christian Science" rather than of, say, "cognitive psychology".

5. It is unclear to what extent earlier versions of RTM accepted this doctrine implicitly. There was, for example, a traditional disagreement over whether association "by similarity" could be a fundamental psychological mechanism, along with association by spatio-temporal contiguity. It is possible to view this debate as in part about whether mental operations are computational. Since the sort of similarity at issue was similarity of *content*, to acknowledge association by similarity as irreducible would have been to violate the condition that mental operations apply in virtue of the *form* of representations in their domains. Contiguity, by contrast, counts as a formal property within the meaning of the act.

6. I shall often use "de dicto propositional attitude" as short for "propositional attitude under de dicto specification". No propositional attitude can be *de dicto per se*, or *de re stricto dictu*. Of course.

7. In a recent paper [8], however, Putnam opts for a "verificationist" notion of *concept* such that "water" and "$water_2$", though nonsynonymous, are nevertheless used to express the *same* concept so long as speakers are ignorant of the relevant microchemical facts. The effect is thus to detach the question "what does the word *W* mean?" from the question "what concept is the word *W* used to express?" and, more generally, to detach semantic questions from questions about the propositional attitudes of speaker/hearers. I think that there is *something* right about this since, as will appear, I think that the content of a belief does not determine its truth conditions (or those of the forms of words used to express it). My reasons for thinking this are, however, a good deal less dramatic than Putnam's. Though it will turn out that beliefs of identical content may differ in their truth conditions, the considerations that lead to this conclusion are philosophically innocuous. They do not, in particular, lend any comfort to verificationists.

8. I propose to use such expressions as "Grice's principle", "Gricean theory", "Gricean reduction", and so forth to refer to any of a wide variety of doctrines which have in common the idea that meanings are somehow constructs out of propositional attitudes. I do not mean to assume the particular account of that reduction which Grice sets out in [6]; nor, of course, is Grice to be blamed for the use I have made of his leading doctrine.

9. This is pointed out by Burge in [1], although the morals he draws are quite different from the ones that I shall endorse.

10. Professor Michael Lipton has suggested to me the following snazzy argument, which makes trouble for all forms of the indexical proposal. What we are trying to do is pick

out the *de dicto* belief that "water$_2$ is wet" is used to express. But you cannot do this with a formula of the form "the belief that this stuff is . . ." because, since indexicals always occur *transparently* in descriptions of propositional attitudes, no such formula can, even in principle, specify a belief *de dicto*. What I suspect that this argument shows is that there can be no word that is defined in terms of an indexical (which is not, of course, to say that there can be no indexical words).

11. Notice that 2 is not a version of Grice's principle, for although it connects semantic properties with properties of concepts, the latter are *not* being assumed by Burge to be mental in anything like the sense that Gricean reductions require.

12. As the reader will have divined, I am pretending that *verbal contracts bind* exhausts the concept of contract instead of just constituting part of it as per assumption 1. It simplifies the discussion and changes nothing to do so.

13. More precisely, I take the following principle to be valid: Let *a* and *b* be distinct expressions, and let *believes that . . . a . . .* and *believes that . . . b . . .* be formulas that specify beliefs *de dicto*. Then, if *a* and *b* are synonyms (express the same concept) either $\Box[(x)$ (x believes that . . . *a* . . .$) \equiv (x$ believes that . . . *b* . . .$)]$ or *believes that . . . a . . .* and *believes that . . . b . . .* designate distinct beliefs.

14. I have no story at all to tell about Burge's examples insofar as they do not involve kind terms, since I find that I do not share many of the intuitions that motivate Burge's solutions. A natural thing to do in the case of "contract" would be to take it to express some such concept as *legally binding agreement*, so that the belief that contracts must be written would be consistent but false. For what it is worth, *The American College Dictionary* says that a contract is "an agreement enforceable by law", thereby leaving it as a legal (rather than a conceptual) issue whether verbal contracts bind. This seems to be entirely plausible, for it seems to me not incoherent to wonder whether, for example, verbal contracts are binding in France.

15. For convenience, I am assuming that "water is wet" expresses an (implicit) universal generalization. But the analysis I shall propose has an obvious extension to the assumption that "water" is a singular term. Indeed, the principles involved in the analysis appear to be quite general in their application. See note 17 below.

16. The "the" is there to indicate that the belief that water is wet purports to be about a *single* kind of stuff, and the "kind" is there to indicate that the belief that water is wet purports to be about a kind. In these respects the present analysis shares Putnam's assumption that "water" is (or, anyhow, purports to be) a kind term.

17. Analogous considerations apply to ensure that the evaluation of existentially quantified variables should also be local, a point that has been widely noticed. For example, nobody would evaluate "there are cookies to eat" with respect to cookies in China. Similarly with respect to singular terms, whose referential ambiguity would otherwise make life miserable. Contemporary tokens of "John is in the shower" are not, of course, evaluated with reference to John the Baptist.

18. Isotropy does not, of course, supply an independent characterization of the relevant universe of discourse since isotropic regions of space-time just *are* those that are homogeneous with respect to the sorts of descriptions that laws deploy.

19. This is not a fudge since, on anybody's story, explicit indexicals can presumably be distinguished from words like "water" on grounds independent of their indexicality. For example, the explicit indexicals belong to "closed class" vocabulary.

REFERENCES

[1] Burge, T., "Individualism and the mental," *Midwest Studies in Philosophy*, vol. 4 (1979), pp. 73-121.

[2] Davidson, D., "Truth and meaning," *Synthese*, vol. 17 (1967), pp. 304-323.

[3] Fodor, J., *Representations*, MIT Press, Cambridge, Massachusetts, 1981.

[4] Geach, P., "Some remarks on representation," *Behavioral and Brain Science*, vol. 3 (1980), pp. 80-81.

[5] Gibson, J., *The Ecological Approach to Visual Perception*, Houghton-Mifflin, Boston, Massachusetts, 1979.

[6] Grice, H., "Meaning," *Philosophical Review*, vol. 66 (1957), pp. 377-388.

[7] Polo, Marco, *The Travels of Marco Polo*, Modern Library, Random House, New York, 1926.

[8] Putnam, H., "Computational psychology and interpretation theory," Harvard University, mimeo.

[9] Putnam, H., "The meaning of 'meaning'," in *Mind, Language and Reality, Philosophical Papers*, Vol. 2, Cambridge University Press, 1975.

[10] Putnam, H., *Meaning and the Moral Sciences*, Routledge and Kegan Paul, Boston, Massachusetts, 1978.

[11] Reid, T., *Essays on the Intellectual Powers of Man*, MIT Press, Cambridge, Massachusetts, 1969 (originally published in 1785).

[12] Stich, S., "Paying the price for methodological solipsism," *Behavioral and Brain Science*, vol. 3 (1980), pp. 97-98.

[13] Stich, S., "Computation without representations," *Behavioral and Brain Science*, vol. 3 (1980), p. 152.

Department of Philosophy
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139