PARTIAL CHARACTER AND THE LANGUAGE OF THOUGHT

BY

STEPHEN L. WHITE

HILARY Putnam has argued in a number of articles that a speaker's psychology fails to determine the content of his beliefs.[1] In a recent series of papers Stephen Stich[2] and Tyler Burge[3] have extended Putnam's suggestions and in so doing have raised what may be among the most serious problems for a functionalist theory of belief. These problems concern the possibility of correlating beliefs and functional states given some assumptions, widely held by causal theorists of reference, about the individuation of beliefs. In the first section I shall examine the arguments advanced by Burge and Stich and discuss some of the difficulties that their own positions raise. In section II I shall consider two plausible strategies for countering their arguments, each of which, as I shall go on to argue, is ultimately unsatisfactory. In section III I propose a more adequate response in defense of functionalism, which is based on the concepts of partial character and context of acquisition. I shall then show in section IV that the concept of partial character is of interest quite apart from its role in a defense of functionalism, and that it plays a part in the analysis of meaning change as well as in our understanding of the notion of a language of thought.

I. Autonomous Psychology and Belief-Desire Explanation

In "Autonomous Psychology and the Belief-Desire Thesis," Stich argues that a tension exists between the following two principles:

(1) Principle of Psychological Autonomy
   The properties and relations to be invoked in an explanatory psychological theory must be supervenient upon the *current, internal physical* properties and relations of organisms (i.e., just those properties that an organism shares with all of its replicas).[4]

(2) Belief-Desire Thesis
   (a) Human action is, at least in part, caused by, and explained in terms of, beliefs and desires.
   (b) Singular causal statements connecting particular beliefs and desires and actions are true in virtue of being subsumed by laws of a psychological theory which specify nomological relations among beliefs, desires, and action.[5]

0031-5621/82/0400-0213$1.90

In other words, a singluar causal statement connecting a belief and an action is subsumed, according to the belief-desire thesis, by a law of a psychological theory which mentions beliefs and desires, not by a law of a physical theory which mentions only physical states, some of which are type or token identical with the beliefs and desires in question.

Stich claims that if the autonomy principle is acepted then there will be a large number of belief properties[6] which could not play a role in an explanatory psychological theory, i.e., there will be a large number of belief properties which will not satisfy the belief-desire thesis. This claim is based on his assumption that a sufficient condition for the nonidentity of two belief properties is that tokens of them differ in truth value. His claim is supported by a number of examples in which an individual has a belief which is true, while the corresponding belief of a physical replica is false. Stich concludes that the individual instantiates a belief property different from the one the replica instantiates and that in view of this difference the principle of psychological autonomy forces the conclusion that the belief property instantiated by the individual is not a property that could figure in an explanatory psychological theory. In other words, no such belief will be subsumed by the laws of an explanatory psychological theory. But, since our belief in the possibility of such a psychological theory rests on our expectation that beliefs of the sort cited in the examples *do* play a role in the explanation of human action, the examples are held to undermine our belief in the possibility of such a theory.

Stich's first example involves indexical beliefs. Stich imagines that a duplicate of himself, in which he is replicated atom for atom, has been created in the last five minutes. Both he and his replica share the belief that would be appropriately expressed by the sentence 'I have tasted a Chateau d'Yquem '62.' But, since his belief is true and his replica's is not, the belief property that he instantiates is not a property shared by all his replicas and thus, by the principle of psychological autonomy, not one that could figure in an explanatory psychological theory. Since the same argument applies to beliefs involving other indexicals such as 'here' and 'now,' indexical beliefs will not occur in the relevant sort of psychological explanation.

Stich's other examples support the same form of argument adding further classes of beliefs which will not figure in psychological explanation. Like indexical beliefs, beliefs involving natural kind terms and proper names can be eliminated from the domain of psychological theory. Following Putnam,[7] we can imagine a planet called Twin Earth which is an almost exact physical replica of Earth, and on which there is an almost exact physical replica of each individual. Twin Earth differs from Earth only in the fact that the natural kind which has all the obvious properties of water and which fills the oceans, lakes and resevoirs is not H_2O but a substance whose formula we may abbreviate XYZ. Putnam's claim is that the expression 'water,' as used by our Twin Earth doubles, refers to XYZ, not to H_2O. Given this assumption, had our ancestors believed that lizards dissolved in water, their beliefs would have been false, while the beliefs which their Twin Earth doubles would have expressed in the same words might have been true—lizards might have dissolved in XYZ. Hence beliefs involving natural kind terms, because of the way such terms pick out their referents, are barred from psychology. A similar argument rules out beliefs involving proper names.

Since Stich's arguments are not explicitly directed at functionalism, it will be useful to pause long enough to make the anti-functionalist implications explicit. I am taking faunctionalism as the thesis that intentional predicates like 'X believes that p' are definable in terms of functional predicates which are themselves defined in terms of the relations between inputs, internal states, and outputs of a system. There are, of course, different versions of functionalism, and the differences will have a bearing on what counts as a functional predicate, but in this context the differences need not concern us. Stich's argument that certain classes of beliefs could not figure in psychological explanations is equally an argument that these beliefs could not have logically necessary and sufficient conditions in terms of functional states. Since one's physical replicas do share all of one's functional states and fail to share many of one's belief states, many different belief states will correspond to the same functional state; hence necessary and sufficient conditions for these belief states could not be given in functional terms. Moreover, the differences between one's own beliefs and the beliefs of one's physical replicas could not get a functional explanation.

Clearly these arguments do not count against all possible versions of functionalism. In particular we might consider an eliminative version of functionalism according to which functional predicates replace rather than define intentional predicates. Stich, in fact, gives some indication that he would favor such a theory. Functional properties do satisfy the principle of psychological autonomy and in many central cases belief properties do not. Hence, in the construction of an explanatory psychological theory it is the belief properties that Stich would have us abandon.[8]

Where Putnam was concerned with the role of the physical environment in determining the meanings of the expressions of a subject, Burge in "Individualism and the Mental"[9] is concerned with the role of the social environment. Burge proposes a thought experiment with parallels to the Twin Earth example which is designed to show that speakers who are functionally and physically indistinguishable may entertain different beliefs, not as a result of proximity to different natural kinds, but as a result of membership in different linguistic communities.

In the first step of the thought experiment Burge imagines a subject whose understanding of the concept of arthritis is incomplete. The subject fails to realize that arthritis is specifically a condition of the joints. Although he has a number of true beliefs about arthritis, for example that he has had arthritis for years, that his arthritis in his wrists and fingers is more painful than his arthritis in his ankles, that stiffening joints is a symptom of arthritis, that certain sorts of aches are characteristic of arthritis, and so forth, he has, in addition to these true beliefs, the false belief that he has arthritis of the thigh.

In the second step of the experiment Burge has us imagine a counterfactual situation in which the speaker remains unchanged, at least physically, while we replace his community with one in which the word 'arthritis' is used to cover various rhumatoid ailments, the one in the subject's thigh as well as the genuinely arthritic conditions.

In the third step Burge asks us to concede that the subject in the actual situation and his alternate self in the counterfactual situation have different beliefs even though they may have been physically indistinguishable throughout their lives. Burge points out that it is hard to see how the subject's alternate self could have

acquired the concept of arthritis and that the concept of arthritis and the concept which was acquired by the subject's alternate self are not even extensionally equivalent. This is because the subject's belief which he would express by saying 'I have arthritis of the thigh' is false, whereas the belief that his alternate self would express in the same way is true. Hence, according to Burge, the subject's alternate self does not have the belief that he has had arthritis for many years, that stiffening of the joints is a symptom of arthritis, and so forth, even though the alternate does have beliefs which he would express in exactly these words— the same words in which the subject himself would express them.

The implications of this example for functionalism are the same as those of Stich's examples, and Burge makes the anti-functionalist consequences explicit. Since the subject and his alternate self share all the same functional states and they do not have the same beliefs, functional states do not explain belief. But, although Burge's example complements Stich's, Burge and Stich draw radically different conclusions. Stich, faced with the incompatibility of the principle of psychological autonomy and the belief-desire thesis, at least insofar as the latter is held to apply to the full range of our belief ascriptions, opts to drop the belief-desire thesis. Burge faced with the same choice chooses to drop the principle of psychological autonomy. Stich's reason for preferring the principle of psychological autonomy is that functional states explain behavior at least under its physical description and, that for the purposes of psychological explanation, this is precisely what we want. Burge, on the other hand, is anxious to emphasize the influence of the social on the mental, as well as to preserve our ordinary forms of belief-desire explanations.

Unlike Burge and Stich, I find the prospect of choosing one of these principles over the other an unattractive one. It is a particularly unpalatable choice for functionalists committed to the idea that functional states explain in some sense our ordinary belief-desire ascriptions. In the next section I shall examine two ways of trying to reconcile the two theses. Although we will see that they are ultimately unsatisfactory, they will shed some light on another more satisfactory method.

II. Stereotypes and Qualitative Predicates

Since the problems that Burge and Stich raise for an autonomous belief-desire psychology are direct applications and extensions of Putnam's claims about reference, the obvious place to look for an answer to these difficulties would be in Putnam's own discussion of meaning and belief. Putnam describes the meaning of a natural kind term in terms of its normal form of description:

My proposal is that the normal form description of the meaning of a word should be a finite sequence, or 'vector,' whose components should certainly include the following (it might be desirable to have other components as well): (1) the syntactic markers that apply to the word, e.g. 'noun'; (2) the semantic markers that apply to the word, e.g. 'animal,' 'period of time'; (3) a description of the additional features of the stereotype if any; (4) a description of the extension.

The following convention is a part of this proposal: the components of the vector all represent a hypothesis about the individual speaker's competence, *except the extension*. Thus the normal form description for 'water' might be, in part:[10]

Syntactic Markers	Semantic Markers	Stereotype	Extension
mass noun, concrete;	natural kind; liquid;	colorless; transparent; tasteless; thirst-quenching; etc.	H_2O (give or take impurities)

This proposal is an attempt to separate the social and the physical contribution to meaning, which is represented in the fourth column, from the individual's contributions, represented in the first three columns. As a representation of meaning, however, this hovers uneasily between two distinct paradigms. Intensions— functions from possible worlds to extensions—determine what is meant without any reference to the rules or recognitional capacities which would give the function psychological reality. The function from possible worlds to truth values associated with 'snow is white,' for example, determines a set of worlds whose only common feature is that snow is white in those worlds, thereby fixing the meaning of the sentence. The conception of meaning which corresponds to intensions, therefore, does not supply a rule, applicable in an indefinite number of circumstances which would allow us to determine whether in those circumstances snow was in fact white. Hence the conception of meaning which *does* correspond to such a rule or recognitional capacity provides a distinct and independent paradigm.

According to Putnam, 'water' picks out H_2O with respect to every possible world. Hence the difference between specifying the intension of 'water' and specifying H_2O is merely the difference between a constant function and the value of that constant function—a difference that can be safely ignored. Thus Putnam's representation of the meaning of 'water' falls under the paradigm associated with intensions. Putnam makes some concession as well, however, to the recognitional capacity conception in his inclusion of the stereotype associated with 'water.' The stereotype does not, of course, fix the referent of 'water' as H_2O, but this is in line with Putnam's contention that what is in the head (the rule or the stereotype) does not fix reference. Still, there are a number of problems which suggest that Putnam has underestimated the extent to which what is in the head *does* contribute to the determination of the referent.

First, there is more to what is in the head than is represented in Putnam's stereotypes. Speakers of the language must have a general commitment to use the language in the same way it is used in the community, and these commitments will be represented by the intentions with which speakers engage in linguistic activity. These intentions (or at least some important aspects of them) will be in the head in Putnam's sense and will help determine the extensions of the speakers' concepts.

Secondly, words like 'elm' and 'beech' which figure prominently in Putnam's examples must be the exception rather than the rule—at least among the words speakers commonly use. A speaker's working vocabulary could not be made up

solely, or even largely, of words which have what we might call a thin stereotype, i.e., a stereotype which does not determine a recognitional capacity which would allow the subject to discriminate typical examples of the concept in question, at least under favorable conditions. If a speaker's working vocabulary were made up of such words, it would be an uninterpreted language as far as he was concerned, and we would cease to regard him as a speaker of that language. We may tolerate speakers who cannot tell elms from beeches, but speakers who cannot associate even such common words as 'chair' with certain typical examples in their environment picked out ostensively will not ordinarily count as speakers of English. And notice that such a person is still in a much stronger position than the one who cannot *tell* elms from beeches since this latter person not only has not associated 'elm' with some class of objects, but lacks a recognitional capacity capable of discriminating typical members of the class under any even roughly coextensional concepts. That is, like the person who cannot distinguish identical twins, he has not recognized those features which make elms and beeches discriminable. Thin stereotypes work because of the division of linguistic labor, but for every meaningful expression some members of the community, the experts, must have thick stereotypes if the expressions are to have an appropriate connection with the world. Moreover, every member of the speech community must have *some* thick stereotypes if he or she is to be capable of communication with the experts.

Thirdly, Putnam never really addresses the problem of semantic compositionality—that is, the problem of how the truth conditions (or whatever properties of sentences come to play the theoretical role ordinarily reserved for truth conditions in a theory of meaning) of complex sentences are generated from their simpler components. Putnam does mention that a speaker may have synonymous terms in his vocabulary which he does not recognize as synonyms, e.g., a bilingual speaker may associate the same stereotype and extension with 'buche' as with 'beech' and believe that this tree is a beech (it has a name plate on it) without believing it is a 'buche' (he is agnostic on this question). In dealing with this case, Putnam introduces the notion of a representation under which the speaker believes a proposition so that the speaker is said to believe the proposition that this is a beech under one representation and not under another. This is just to say, presumably, that the truth conditions of some belief sentences depend, in addition to their dependence on the truth conditions of the embedded proposition, on the way in which the believer represents the proposition to himself. Thus,

(1) Oscar believes that this is a beech

will be true and

(2) Oscar believes that this is a *buche*

will be false even though

(3) This is a beech

and

(4) This is a *buche*

express the same proposition. Putnam seems to be suggesting that at least certain important features of the respective logical forms are represented by

(5) Believes [Oscar, P, R_1]
(6) Believes [Oscar, P, R_2]

where P is the proposition expressed by (3) and (4) and R_1 is the representation under which Oscar believes it and R_2 the representation under which he is agnostic. Thus we can say again that the extensions of some of a speaker's utterances such as

(7) I believe this is a beech

and

(8) I believe this is a *buche*

depend on more than his thin stereotypes. But Putnam says very little about the relation between stereotypes and representations.

Finally, another aspect of meaning which is, in one clear sense, in the head, and which helps determine the extensions of our concepts is the set of dispositions that govern the similarity judgments which we make in novel situations. Here, too, the concepts that an individual possesses are not completely settled by what is in *his* head, but neither are they settled by experts. The boundaries of predicates such as 'is red' or 'is funny' are settled at least partly by the extra-linguistic dispositions to make similarity judgments of a majority of the speech community.

What is being criticized is the lack of an adequate theory of the way in which what is in the head contributes to the determination of meaning and reference. Putnam does provide a way of separating the individual and external components of meaning, but he provides no mechanism for the connection between the internal component of the meaning of a linguistic expression and the extension. The best indication of this is the fact that we can imagine people with precisely the internal components of meaning postulated by Putnam and living in a world with precisely the natural kinds that our world has, but who have concepts with different extensions (and therefore, in Putnam's sense, different meanings). In the first place, a group of people who had vocabularies composed of only those words for which we have, according to Putnam, thin stereotypes would probably not count as speakers of a language at all—even if the group consisted of quasi-experts who did associate detailed definitions couched in terms of other words in the vocabulary with a selected subset of the words. Similarly, a group of people who had different dispositions regarding similarity judgments would associate different intensions and extensions with many of their words, even if they shared all of our thin stereotypes. Furthermore, people who lacked our linguistic intentions to use words in accordance with the use of the community and the experts would also lack many of our extensions.

The upshop of this is that a theory which effected a division of the internal and external components of meaning without these defects would clearly be welcome. In fact, Putnam's remarks on the indexicality of terms like 'water' suggest a different approach to the internal and external contributions to meaning.

Now then, we have maintained that indexicality extends beyond the *obviously* indexical words and morphemes (e.g. the tenses of verbs). Our theory can be summarized as saying that words like 'water' have an unnoticed indexical component: 'water' is stuff that bears a certain similarity relation to the water *around here*.[11]

The idea that this remark suggests (but does not endorse) would take Putnam's own characterization of water, "stuff that bears a certain similarity relation to the water *around here,*" as the basis of a theory which we might call the indexical-description theory of meaning. The meaning of 'water' would be characterized by descriptive terms, for example, 'whatever is wet, fills the lakes, comes out of the faucets, etc.,' which would pick out water in the actual world and which, together with the explicit indexical 'here,' would insure that 'water' as used by us functions rigidly to denote, in every possible world, H_2O. 'Water' as used on Twin Earth will have a different meaning, at least in the sense that it will denote XYZ in any possible world.

The indexical-description theory is an attempt to address simultaneously two problems. The first problem is just the problem of effecting the division of internal and external components of meaning already mentioned. This is accomplished by isolating the semantic differences between speakers on Earth and their Twin Earth counterparts in the explicitly indexical portions of their respective languages and locating the semantic similarities in the descriptive portions of those languages. The second problem involves accounting for the determination of the speaker's extensions by the internal components of meaning, and this is handled by invoking the descriptions under which the speaker apprehends the referents of his expressions while permitting the description to contain explicit indexicals to account for rigid designations and "circular" descriptions to allow for cases in which the speaker succeeds in referring even without access to a noncircular description which individuates the object of reference.

Not only does this theory purport to solve both problems, but it seems to provide the two problems with a common solution which is supported by two kinds of considerations. First, consideration of what the physical duplicates share semantically suggests that even if they associate *different* extensions with words such as 'water' this must be in virtue of sharing predicates with which they do associate the same intensions and extensions, e.g., 'wet,' 'thirst-quenching,' etc. Call such predicates the autonomous predicates. Their intensions and extensions do not depend on environmental facts which could vary from one physical duplicate to another, hence they must be applicable on the basis of features which are fully manifest to each subject. The intuition is that what is in the head determines the intensions and extensions of the autonomous predicates, and these together with explicit indexicals determine the intensions and extensions of other predicates and singular terms. Thus by providing relatively rich descriptions in terms of autonomous predicates and explicit indexicals the theory provides a means by which what is in the head determines (together with the external contributions) the intensions and extensions. Secondly, the theory eliminates the earlier counterexamples (e.g., the example of the "speakers" whose language is uninterpreted), specifies the semantic properties that the functional duplicates share, and provides a plausible approach to semantic compositionality since we allow

the descriptions to function as the meanings of singular terms and not only as a means of fixing their referents.

In spite of these advantages, advocates of such an approach face a difficult dilemma. Either there is no reason to believe in the existence of the autonomous predicates postulated, or they will turn out to be essentially private and therefore inappropriate candidates for the role specified for them by the theory. The argument for this contention is quite straightforward. Unless the autonomous predicates are themselves capable of acquiring different intensions and extensions in different physical environments, their stability will be purchased at the cost of making them inappropriate vehicles of intersubjective communication. Take 'is wet.' Though apparently qualitative, wetness on Twin Earth might well be a property not of liquids but of quantities of very small solid particles which slide by one another easily. Thus the substances which exhibit the macro-properties of liquids on earth are on Twin Earth more like quantities of very fine sand or graphite particles. Even if we move toward predicates like 'looks red to me,' the same problem arises. Situations in which earthlings would be confident in saying that an object looked red to them might be situations in which the object was in fact invisible but emitted ultrasonic sound waves which produced subliminally, by hypnotic suggestion, the image of a red object with the features of the actual object.

In the case of 'looks red to me' the difficulty is engendered in part by certain features of the word 'looks,' but the word 'red' raises equally serious problems. Imagine a case in which a physical duplicate of a normal speaker of English lives in a community in which the word 'red' is applied to a wider band of the spectrum than it is among speakers of English. In such a case, if Burge is right, the duplicate's concept of red would not be ours and even the retreat to 'seems red to me' could not produce an autonomous predicate in the sense defined above. Apparently only the complete retreat to pure sense data, assuming that such a retreat is intelligible, could rule out such difficulties, and such a retreat seems to preclude the straightforward explication of our communicative abilities which the indexical-description theory was intended to produce.

These arguments do not prove, nor is it necessary to argue, that no version of the indexical-description theory is viable. They do show that none of the *usual* strategies for effecting a division of our vocabulary (or of the concepts definable in terms of our vocabulary) into qualitative and nonqualitative concepts will succeed, and thus that there is no very compelling reason to believe in the existence of autonomous concepts. Thus a theory which does not presuppose the possibility of such a division would be a far more attractive solution to the problem of separating the individual and social contributions to the intensions and extensions of the speaker's concepts.

III. Partial Character

I have claimed that the indexical-description theory attempts to separate the individual and external contributions to the intensions and extensions of a speaker's expressions by dividing the speaker's vocabulary. Explicit indexicals are the locus of both the individual and external contributions, autonomous predicates

are the locus of the individual contributions, and other terms, singular and general, have senses which are identical with the senses of descriptions made up of autonomous predicates and explicit indexicals.

The alternative approach that I want to suggest does not locate the indexical component, and thereby the external contribution, in a subset of the speaker's vocabulary, but spreads both the individual and the external contributions over the entire vocabulary. The result is that each word or concept contains both an indexical and an autonomous component. Since this is true of the explicit indexicals on the indexical-description theory, we will need a counterpart to the distinction, which is recognized regarding explicit indexicals, between those aspects of meaning which are shared by all speakers and those aspects which vary with the differences in the speaker's circumstances.

The clearest characterization of these different aspects of indexicals is found in David Kaplan's distinction between content and character.[12] Kaplan's characterization depends on distinguishing *circumstances* at which an utterance may be evaluated and *contexts* in which an utterance may be used. Since I intend to draw a distinction which Kaplan does not make, I shall speak instead of contexts of evaluation and contexts of utterance.

Kaplan's distinction between content and character is clearest in the cases of sentences like

(1) I am not speaking now.

Clearly (1) is false whenever it is uttered, but is is not necessarily false. On any occasion on which it is uttered the speaker might not have uttered it. Thus there must be possible worlds in which it is true. Finding these worlds, of course, could not require that we look for situations in which someone uttering the sentence is not speaking. Instead we take an utterance of the sentence in some context of utterance, fix the speaker, and look at the cotemporal activities of that person in other possible worlds or contexts of evaluation. Those in which he is not speaking are the worlds in which the sentence is true relative to the context of utterance chosen.

The choice of a context of utterance determines a function from contexts of evaluation to truth values. Kaplan calls such a function a content. Ordinarily it is called an intension, and this is the term I shall use throughout the paper. 'Content' will be used only in a strictly pre-theoretical sense. Since there is nothing special about the actual context of utterance, we need another function from contexts of utterance to intensions, i.e., a function from contexts to a function from contexts to truth values. Kaplan calls such a function the character of an utterance, and in this case I shall adopt his terminology.[13]

For Kaplan, a variable character (i.e., a nonconstant function from contexts of utterance to intensions) is associated with indexical expressions such as 'I,' 'now,' 'you' and so forth. Nonindexicals have a constant character since their intension is a constant function of their context of utterance; every context of utterance is mapped into the same intension. The character of complex expressions is a function of the character of their constituent components.

Putnam's claim, supported by the Twin Earth example, amounts to the thesis that the intension and the extension of even such ostensibly nonindexical words

as 'water' (or, more accurately, utterances of those words) vary according to the context in which the word was acquired. But any proposal to define a function from the contexts of acquisition of words to the characters associated with the words, by analogy with the function from contexts of utterance to intensions, faces the difficulty that nothing will distinguish the words except their physical shape. Whereas 'I' retains its reference to the speaker in any context of utterance, if we allow the context of acquisition to vary, and this includes the training involved in the word's use, 'red' could mean anything from *green* to *garrulous*. Thus the words will have no fixed significance which remains invariant from one context of acquisition to another analogous to the character which 'I' retains in different contexts of utterance.

Rather than allowing the contexts of acquisition of a word to vary at random, I propose to look at that subset of the set of possible contexts of acquisition in which speakers who have acquired the word in question are physical and/or functional duplicates. Let us define the partial character of a word W as follows:

A *possible context of acquisition* is an ordered pair consisting of a possible world and a functional state token which exists in that world. We say that the possible world is *centered on* the token. (Intuitively the possible world fixes the causal history and the causal environment of the token.)

Two possible contexts of acquisition are *functionally equivalent* if and only if they are centered on functional states which are type identical (or "sufficiently close").

Possible context of acquisition CA = ⟨the actual world, functional state token S⟩ is a *context of acquisition of a word type W* if and only if S is the functional state token underlying a use of W. For CA = ⟨possible world PW, functional state token S⟩, CA is a *possible context of acquisition of a word type W* if and only if CA is functionally equivalent to a context of acquisition of W.

An *interpretation* of a word type W is an element of the set E of equivalence classes of possible contexts of acquisition of W generated by the relation of functional equivalence.

A possible context of acquisition CA = ⟨possible world PW, functional state token S⟩ of word type W *fixes the character of W as C* if and only if, given the causal relations between S and its social and physical environment in PW, a correct causal theory of reference entails that the character of W is C.

The *partial character* of a word type W relative to an interpretation I of W is a function F from I to the set of characters such that for CA ∈ I, F(CA) = the character C such that CA fixes the character of W as C.

The easiest way of seeing the definitions at work is by applying them to the Burge and Stich examples. Consider Stich's example in which he and his double both believe that they have tasted a Chateau d'Yquem '62. Stich's belief is true. His double, having been created in the past five minutes, has a false belief. Therefore, they have different beliefs in spite of being in the same functional state. The semantic property which Stich and his double share is that their sentences have the same character. Since their contexts of utterance (or more properly their contexts of occurrence) differ, however, the same character maps their utterances into different intensions and in this case into different extensions. It is clear, however, that in this kind of case and for the purposes of psychology,

character is the relevant dimension along which beliefs are to be assessed. As John Perry has pointed out,[14] if Smith believes he is being attacked by a bear, and Jones shares this belief in the sense that his belief has the same intension as Smith's, then whereas Smith will curl up in a ball, Jones will run for help. If, on the other hand, Jones shares Smith's belief in the sense that his belief has the same character as Smith's, Jones will curl up in a ball as well. The same remarks apply to Stich's second case involving implicit indexical references to a time and location.

It is in the case of proper names and natural kind terms that the distinction between character and partial character comes into play. Stich believes that Saul Kripke was born in Nebraska and Stich's double on Twin Earth has the same belief, though his belief is about Saul Kripke's double. Here we can say that though their beliefs differ in character, intension and extension, they share the same partial character. If Stich had acquired his belief on Twin Earth in the normal way, then even expressed on earth the belief would have been about Saul Kripke's double and would have been true at those possible worlds in which Saul Kripke's double is born in the Twin Earth counterpart of Nebraska. Stich's H_2O and XYZ example is handled by the same analysis, and Burge's arthritis example is also amenable to this kind of treatment. If Burge's counterfactual counterpart of his hypothetical user of the term 'arthritis' had acquired the word in a context like ours, then like the hypothetical speaker he would have had a false belief about arthritis and not a true belief about some more general rhumatoid condition.

It is possible, then, to concede to Burge and Stich that there is a clear sense in which functional duplicates do not have the same beliefs, hence a clear sense in which a subject's functional states do not explain his beliefs. The crucial point is that there is an equally clear sense in which functional duplicates do share the same beliefs; they share the same partial characters. The burden of the rest of the paper will be to show that to share the same partial character *is* in one interesting sense to share the same belief. Moreover, I shall show that partial character is a notion of general theoretical interest quite apart from its role in the issues raised by Burge and Stich.

One criticism of this treatment, which will quickly occur to anyone sympathetic to either the Stich or Burge approach, would be the claim that the functional and/or physical similarity which, by hypothesis, the speakers share with their doubles has merely been renamed. Part of the answer to this criticism will come in the application of the notion of partial character. The first and most important point to be made, however, is this: partial character is a semantic notion and one which allows us to describe those respects in which functionally similar speakers share the same semantic properties. To this it might be objected that although there is a point to recording the semantic similarities of those speakers whose beliefs relate them to the same intensions, there is no corresponding point in attempting to catalogue the semantic similarities of speakers whose beliefs do not relate them to the same intensions (putting aside for the moment the case of explicit indexicals and characters). In the latter case, it might be maintained, no theoretical advantage is to be gained by postulating similarities over and above the functional similarities.

Let us call a theory of semantic content which has this implication a restrictive

theory of semantic ascription. One strong motivation for ascribing semantic similarities over and above functional similarities stems from the problems presented by nondenoting terms for a causal theory of reference. The notion of partial character allows us to reconcile Kripke's claim that if fossils of a species similar to horses with a single horn were discovered they would not be unicorns[15] with the obvious truth that not all nondenoting terms have the same meaning. If Kripke is right 'unicorn' and 'griffin' pick out the same extension at every possible world—the null set. Hence the possible world apparatus will not be sufficient to distinguish them in meaning since they are both associated with the same intension. When we look at partial characters, however, the picture changes significantly. Consider a possible world in which creatures—call them U and G—satisfying the descriptions associated with 'unicorn' and 'griffin' exist, in which those of our ancestors who believed in unicorns and griffins have alternate selves who are functional (and physical) duplicates, and in which the beliefs of their alternate selves are in the right sort of causal contact with these creatures. In such a possible world—call it PW—a typical inhabitant will associate a different intension with 'unicorn' from the one which he associates with 'griffin.' But this means that his alternate self in the actual world will associate a different *partial character* with 'unicorn' from the one he associates with 'griffin' since for a believer of the myths in the actual world, the functional states associated with 'unicorn' and 'griffin'—call them S_1 and S_2—will be such that *if* he had acquired them in PW, his state S_1 *would* have been associated with the intension taking a possible world into the set of U's in that world, and his state S_2 would have been associated with the intension taking a possible world into the set of G's in that world.

Such cases of nondenoting terms have a natural generalization in the case of the brain in the vat. Suppose we imagine a brain which is a functional duplicate of a normal psychological subject and which is being fed completely fabricated sensory stimuli by a group of neurophysiologists. We then have a case in which the causal theory of reference provides no guidance in ascribing semantic content to the functional states in question but in which the subject's experience viewed from the inside is identical to that of a normal subject. Hence there ought to be the same utility in going beyond functional ascription to semantic ascriptions in the case of the brain in the vat as there is in the case of a normal subject—at least if our goal is an understanding of the brain's psychology or of the brain's "world." Partial character allows such a semantic characterization.

IV. The Application of Partial Character

We have now seen that the concept of partial character is capable of performing the function for which it was designed. It provides a way of treating the semantic similarities between functional duplicates in diverse environments, reconciles autonomous psychology with the belief-desire thesis and reconciles functionalism with the indexicality of many ordinary predicates. There are, moreover, less obvious contexts in which a slight generalization of the notion of partial character may be used to advantage. Consider Hartry Field's example of the term 'mass'

in Newton's idiolect.[16] Newton believes many things of mass and roughly speaking, according to Field, half are true of rest mass and half are true of relativistic mass. The principle of charity, which advises us to maximize the number of true beliefs of the subject whose language is being interpreted, provides no guidance here. Either translation will result in the same number of true beliefs. Nor could we maintain that in Newton's idiolect 'mass' lacked reference since, according to Field, so many of Newton's beliefs that were obviously true would turn out to be false. Field claims that it is genuinely indeterminate whether Newton's 'mass' referred to rest mass or relativistic mass.

As it stands the notion of partial character is no help in this situation. Partial character maps contexts of acquisition into characters which determine intensions; hence, partial character explains how what is in a speaker's head can determine different intensions in different contexts. Since Newton's meaning is indeterminate even after we have been *given* all the information about the context of acquisition, partial character has no immediate relevance.

Instead of taking partial character as a mapping (from contexts of acquisition to characters) which determines particular intensions when the characters are constant functions (as they are for nonindexicals), let partial character be a mapping from contexts of acquisition to sets of characters, each member of which (for nonindexicals) determines a single intension. Such a set of characters I shall call a *determined character set*. Hence the context of acquisiton determines a set of intensions which may or may not be a unit set. In other words, some contexts of acquisition narrow the meaning of a particular term down to a single intension and others do not. Newton's context of acquisition, because it led to his observation of both rest mass and relativistic mass, did not determine a single intension.

Field's solution to this problem has definite parallels with this treatment in terms of partial character . Field does not define partial denotation or signification, but intuitively the idea is as follows. Certain considerations about a speaker's causal relations to extra-linguistic objects and his dispositions to assent to sentences ordinarily allow us to say that his singular terms denote certain objects and that his predicates have certain significations (extensions). Sometimes, however, these considerations will fail to determine a unique object as the referent of a singular term or a unique set as the signification of a predicate. In such cases each of the objects or sets will be a partial denotation or signification of the term in question.

Field's references to the causal relation between the speaker and objects and his reference to the speaker's dispositions correspond *roughly* to my notion of the external and internal contributions to meaning, though Field does not consider that particular part of the external contributions made up of the dispositions and causal relations to extra-linguistic reality of the other members of the speaker's community. Nor does Field dissociate the two contributions to meaning as I have and hence cannot show the variation in the meanings of words which corresponds to the variation in the context of acquisition while the functional make-up of the speaker is held constant. Furthermore, Field avoids possible world semantics by assigning extensions rather than intensions and does not distinguish intensions and character. Thus the analogue to Field's notions of partial denotation and partial significance is the idea that a given context of acquisition may only narrow

the partial character of a term down to a set of intensions rather than to a unique intension.

Field's approach has come under criticism from those who believe that his presentation of the facts of the Newton case is faulty and that the actual cases of indeterminacy found in the history of science are of an entirely different sort. Philip Kitcher has recently offered a somewhat different analysis of the kinds of cases Field is concerned with and suggested that his analysis does better justice to the actual cases in the history of science than Field's does.[17] As we have seen, Field associates a set of objects or a set of sets of objects with expression types. It is said that the expression type partially signifies each object or set of objects. Kitcher, on the other hand, associates a set of initiating *events* with an expression type, i.e., events which initiate the use of a token of the expression type, and each initiating event is associated with a single object or extension. Thus, in general, each token of the expression type will be associated with only *one* object or set of objects. Only those token expressions whose utterance is related in the right way to more than one initiating event will be associated with more than one extension, and Kitcher seems to assume these cases will prove the exception rather than the rule.

It seems clear that Kitcher and Field are operating with two distinct paradigms of the sort of indeterminacy of meaning that conceptual change in the history of science generates. Kitcher's paradigm is perhaps best understood in terms of the nonscientific example he provides. Kitcher imagines a famous and eccentric millionairess, Eustacia Evergreen, who, in an effort to secure a measure of privacy, hires an imposter to lead her public life in a new community. Residents of the community who had produced tokens of 'Eustacia Evergreen' as a causal result of newspaper reports will eventually come to produce such tokens as a causal result of contact with the imposter as well. Kitcher claims that those tokens which are the right sort of causal result of the newspaper reports will refer to Eustacia, those which are the right sort of result of contact with the imposter will refer to the imposter, and that the reference of some tokens may be indeterminate. In general, however, Kitcher's assumption is that most (or many) tokens will be traced back to a single initiating event which will involve one referent or the other but not both.

Field does not provide a nonscientific example of the sort of indeterminacy he has in mind, but the construction of one comparable to the Eustacia Evergreen example will help reveal the differences in the cases with which Field and Kitcher are concerned. Imagine a mountainous region in which all and only tall people are beautiful. The inhabitants of the region have only one word, 'beautiful,' for the people who are both tall and beautiful. They also hold certain theoretical views about "beauty" based on a rudimentary theory of genetics; they believe that "beauty" is inherited from the mother's side of the family and that it skips a generation.

Now let us suppose that in fact height is inherited from the mother's side of the family and that beauty is ordinarily passed on from either side. However, at the altitudes these mountain dwellers inhabit the mechanism whereby the father's side of the family passes on beauty is inhibited, and the net effect is that both height and beauty are passed on from the mother's side alone. Beauty, on the

other hand, does skip a generation and height does not. But at the very high altitudes there is a one-generation delay in the mechanism whereby height is passed one. Again the net effect is that beauty and height always coincide.

If we ask whether the native term'beauty' should be translated in our language as 'beauty' or 'height' the question seems genuinely indeterminate. Here the difference between the Field paradigm and the Kitcher paradigm will become apparent. Kitcher's technique for resolving indeterminacy in translation involves reference to a set of initiating events which Kitcher seems to assume will involve at most one of the entities or properties in question. Field's discussion, by contrast, concerns cases in which all the events which might be regarded as initiating events involve both (or all) of the entitiies or properties, reference to which we might be tempted to attribute to the subject. Thus it seems clear that the Kitcher strategy is not an alternative to Field's over the complete range of possible examples and that the Eustacia Evergreen example is not an adequate paradigm for all the cases of conceptual change in science that we might imagine. Thus Kitcher's analysis is best regarded as being complementary to Field's, though it may, as he claims, provide a better analysis of the *actual* cases of alleged indeterminacy in the history of science.

In spite of the differences in these two analyses, the concepts of partial character and context of acquisition can be used, with a slight modification, to express Kitcher's proposal as well as Field's. The context of acquisition of 'Eustacia Evergreen' is an ordered pair consisting of the actual world and a functional state token underlying a use of 'Eustacia Evergreen.' Hence the context of acquisition includes the entire casual history of the token. We can, however, distinguish two subcontexts of the context of acquisition. That is, we can look at two subsets of the set of causal chains relevant to the production of the functional state token underlying 'Eustacia Evergreen': those chains which involve the real Eustacia Evergreen such as the newspaper reports and those which involve the imposter.

Since the whole context of acquisition involves both Eustacia Evergreen *and* the imposter, the value of the partial character at this whole context of acquisition is a determined character set, that is, a set of two characters, containing one function (character) corresponding to Eustacia Evergreen, i.e., a constant function (character) from contexts of utterance to a function (intension) from possible worlds to Eustacia Evergreen, and another function (character) corresponding to the imposter. Because there are distinguishable subcontexts, however, we can use a modified notion of the character of the expression type to associate *different* intensions with different tokens in different contexts of utterance. We simply specify that for expression types and their underlying functional states whose contexts of acquisition break down into subcontexts, the *modified character* of the expression type maps contexts of utterance in which the utterance *is caused by* one subcontext of acquisition into the intension which would have been the only intension associated with the expression type if the subcontext had been the only part of the whole context relevant to the acquisition of the functional state. Hence we have one context of acquisition CA_1, but two subcontexts SCA_1 and SCA_2, defined by two different initiating events (or types of events), e_1 and e_2. Let the partial character on the Kitcher analysis PC_K take that *one* context of acquisition CA_1 into a determined character set DC_K consisting of *one* function (modified character) C_K which maps contexts of utterance in which the token

utterance was caused by Eustacia Evergreen into Eustacia Evergreen at each possible world and contexts of utterance in which the token utterance was caused by the imposter into the imposter (at each possible world). Thus the notion of partial character allows the construction of a unified framework in which the two approaches to indeterminancy represented by Field and Kitcher may be represented. Examples illustrating the differences between Field's approach and Kitcher's by reference to partial character are fully worked out in the appendix.

As a final instance of the relevance of partial character to philosophy of language consider the problem of reconciling the desire for a semantics of the language of thought with the causal theory of reference. The fact that meaning is nonautonomous for the causal theorist has led at least one such theorist to despair of the possibility of explaining what it is to believe that Caesar was eqotistical without relying at any point on the semantic features of the sentence 'Caesar was egotistical' in one's spoken or written language.[18]

My proposal is that the semantics of an internal system of representations can be done independently of the semantics of a subject's natural language *without* giving up the causal theory of reference. We can do the semantics of the internal language by assigning partial characters to expressions rather than by characterizing the expressions in terms of such nonautonomous properties as intensions and extensions. The semantics of the internal system of representations would be independent of the *actual* semantic features of the natural language though not of its potential semantic features. Another way of putting this is to say that the compatibility of internal semantics and the causal theory of reference is possible precisely because the internal semantics can invoke potential causes to explain potential references. And, of course, the semantic features of the natural language will not be *reducible* to the semantic features of the internal system of representation. The notion of partial character allows us to make sense of the notion of a semantics of a language of thought and hence of the notion of a language of thought itself.[19]

The University of Michigan
Ann Arbor, Michigan

Appendix

Example 1. The Eustacia Evergreen example, not incorporating Kitcher's analysis of utterance tokens.

    CE = context of evaluation
    CU = context of utterance
    CA = context of acquisition
    PW = possible world
    EE = Eustacia Evergreen
    Im = the imposter
  {CE} = the set of possible contexts of evaluation
  {CU} = the set of possible contexts of utterance
  {CA} = the set of possible contexts of acquisition (relative to a fixed interpretation)
   {E} = the set of possible extensions
   {I} = the set of possible intensions

$\{DC\}$ = the set of possible determined characters
I = intension
C = character
DC = determined character set
PC = partial character
$I : \{CE\} \rightarrow \{E\}$
$C : \{CU\} \rightarrow \{I\}$
$DC = \{C_1, \ldots, C_n\}$
$PC : \{CA\} \rightarrow \{DC\}$

Let
$I_1 = \{\langle PW_1, EE \rangle, \langle PW_2, EE \rangle, \langle PW_3, EE \rangle, \ldots\}$
$I_2 = \{\langle PW_1, Im \rangle, \langle PW_2, Im \rangle, \langle PW_3, Im \rangle, \ldots\}$
$C_1 = \{\langle CU_1, I_1 \rangle, \langle CU_2, I_1 \rangle, \langle CU_3, I_1 \rangle, \ldots\}$
$C_2 = \{\langle CU_1, I_2 \rangle, \langle CU_2, I_2 \rangle, \langle CU_3, I_2 \rangle, \ldots\}$
CA_1 = the whole *context of acquisition* for the exression type 'Eustacia Evergreen', i.e., the context of acquisition in the actual world which includes the newspaper reports and meetings with the imposter.
$DC(CA_1)$ = the *determined character set* of the expression type 'Eustacia Evergreen' for the argument consisting of the whole context of acquisition = $\{C_1, C_2\}$.
PC = the *partial character* of the expression type 'Eustacia Evergreen' = $\{\langle CA_1, \{C_1, C_2\}\rangle, \ldots\}$.

Example 2. The Eustacia Evergreen example incorporating Kitcher's analysis.

Let
I_1 = same as Example 1
I_2 = same as Example 1
SCA_1 = the subcontext of the whole context of acquisition in which Eustacia Evergreen figures.
SCA_2 = the subcontext of the whole context of acquisition in which the imposter figures.
CU_1 = a context of utterance in which the token utterance of 'Eustacia Evergreen' is caused by the newspaper reports about Eustacia Evergreen.
CU_2 = a context of utterance in which the token utterance of 'Eustacia Evergreen' is caused by a meeting with the imposter.
C_K = the *modified character* of the expression type 'Eustacia Evergreen' = $\{\langle CU_1, I_1 \rangle, \langle CU_2, I_2 \rangle, \ldots\}$.
$DC_K(CA_1)$ = the *determined character set* (on the Kitcher analysis) of the expression type 'Eustacia Evergreen' for the argument consisting of the whole context of acquisition = $\{C_K\}$.
PC_K = the *partial character* (on the Kitcher analysis) of the expression type 'Eustacia Evergreen' = $\{\langle CA_1, \{C_K\}\rangle, \ldots\}$.

NOTES

[1]Hilary Putnam, "Is Semantics Possible?," reprinted in *Mind, Language and Reality* (Cambridge: Cambridge Univ. Press, 1975), pp. 139–152; "Explanation and Reference," *Mind, Language and Reality*, pp. 196–214; "The Meaning of 'Meaning'," *Mind, Language and Reality*, pp. 215–271; "Meaning and Reference," *The Journal of Philosophy* 70 (1973), 699–711.

[2]Stephen Stich, "Do Animals Have Beliefs?" *Australasian Journal of Philosophy* 57 (1979), 15–28; "Autonomous Psychology and the Belief-Desire Thesis," *Monist* 61 (1978), 573–591; "On the Ascription of Content," forthcoming.

[3]Tyler Burge, "Individualism and the Mental," *Midwest Studies in Philosophy* 4 (1979), 73–121.

[4]Stich, "Autonomous Psychology and the Belief-Desire Thesis," p. 575.

[5]*Ibid.*, pp. 576–578.

[6]Stich takes belief tokens to be states of a person. States are taken to be the instantiation of properties by objects during time intervals. Two belief states (token beliefs) are of the same type if they instantiate the same property and are of different types if they instantiate different properties. If two people believe that power corrupts, then the property they both instantiate is *believing that power corrupts*.

[7]Putnam, "The Meaning of 'Meaning'," pp. 223–227.

[8]Stich, "Autonomous Psychology and the Belief-Desire Thesis," pp. 585–586.

[9]Burge, op. cit., pp. 73–121.

[10]Putnam, "The Meaning of 'Meaning'," p. 269.

[11]*Ibid.*, p. 234.

[12]David Kaplan, "Demonstratives," mimeographed 1977.

[13]This is a slight oversimplification. For Kaplan an intension is not *identical* to the corresponding content, it merely *represents* that content. Similarly the function from contexts of utterance to contents *represents* the character. I shall continue to ignore this distinction since Kaplan himself ignores it in most contexts and since it is irrelevant to the points made in this paper.

[14]John Perry, "Frege on Demonstratives," *The Philosophical Review* 86 (1977), 494.

[15]Saul Kripke, *Naming and Necessity* (Cambridge: Harvard University Press, 1980), pp. 23–24.

[16]Hartry Field, "Conventionalism and Instrumentalism in Semantics," *Nous* 9 (1975), 375–405. "Quine and the Correspondence Theory," *The Philosophical Review* 83 (1974), 200–228.

[17]Philip Kitcher, "Theories, Theorists and Theoretical Change," *The Philosophical Review* 87 (1978), 519–547.

[18]Hartry Field, "Mental Representation," *Erkenntnis* 13 (1978).

[19]I am grateful to Kit Fine, Jaegwon Kim, William Lycan, Timothy McCarthy, Stephen Schiffer, and Nicholas White for comments and advice on earlier drafts of this paper.