MIDWEST STUDIES IN PHILOSOPHY, X (1986)

Advertisement for a Semantics for Psychology

NED BLOCK

Meaning is notoriously vague. So, it should not be surprising that semanticists (those who study meaning) have had somewhat different purposes in mind, and thus have sharpened the ordinary concept of meaning in somewhat different ways. It is a curious and unfortunate fact that semanticists typically tell us little about what aspects of meaning they are and are not attempting to deal with. One is given little guidance as to what extent "rival" research programs actually disagree.

My purpose here is to advocate an approach to semantics relevant to the foundations of psychology, or, rather, one approach to one branch of psychology, namely cognitive science. I shall be talking in terms of some of the leading ideas of cognitive science, most importantly the representational theory of mind, aspects of which will be sketched as they become relevant.[1] The representalist doctrine that my argument depends on is that thoughts are structured entities. I know this will be a sticking point for some readers, so I will say a bit more about what this comes to, and I will compare my position with related positions that reject it.

My strategy will be to begin with some desiderata. These desiderata vary along many dimensions: how central they are to meaning, how psychologically oriented they are, how controversial they are. I will argue that one approach to semantics (not to keep you in suspense—conceptual role semantics) promises to handle such desiderata better than the others that I know about. Though I think my desiderata yield a coherent picture of a psychologically relevant semantics, they are not intended to be pretheoretically obvious; rather, they were chosen to flatter the theory I have in mind. I will *not* be arguing that semantic theories that fail to satisfy these desiderata are thereby defective; there are distinct—and equally legitimate—questions about meaning that a semantic theory can seek to answer.

The view that I am advertising is a variant on the functionalism familiar in the philosophy of mind. However, I will not be attempting to counter the objections that have been raised to that view (except briefly, and in passing). My bet is that looking at functionalism from the point of view of meaning (rather than mentality) and with an eye to its fertility and power rather than its weaknesses will provide a rationale for working on its problems.

DESIDERATA

Desideratum 1: Explain the relation between meaning and reference/truth. This is the least psychological of all my desiderata. The details of what I have in mind will be discussed when I say how conceptual role semantics promises to explain the relation between meaning and truth.

Desideratum 2: Explain what makes meaningful expressions meaningful. What is it about 'cat' in virtue of which it has the meaning it has? What is the difference between 'cat' and 'glurg' in virtue of which the former has meaning and the latter does not? (And so on, for types of expressions other than words.)

Desideratum 3: Explain the relativity of meaning to representational system. This desideratum is arguably just a special case of the preceding one, but I think it is worth mentioning and discussing separately. As we all know, one linguistic item—for example, a sound or linguistic expression—can have different meanings in different languages. For example, many vocabulary items have different meanings in the dialects of English spoken in North America and England, as in 'trailer' and 'bathroom'.

But the significance of this relativity of meaning to system of representation goes deeper than such examples suggest. One way to see this is to note that whole semantic (and syntactic) *categories* are relative to system of representation. Ink marks that function as a picture in your tribe may function as a word in mine. Further, within the category of pictures, representations are understood differently in different cultures.[2] Finally, syntactic category is relative in the same way. Handwriting, for example, differs in different school systems. Perhaps the ink marks that are regarded as an 'A' in Edinburgh are regarded as an 'H' in Chicago. Is there some common explanation of the relativity to representational system of both semantic and syntactic categories?

Desideratum 4: Explain compositionality. The meaning of a sentence is in some sense a function of the meanings of the words in it (plus the syntax of the sentence). What, exactly, is the relation between the semantic values of sentences and words? Is one more basic than the other? Another question arises once we have fixed on an answer to these questions—namely, why is

it that the semantic value of a sentence has whatever relation it has to the semantic values of its parts?

Desideratum 5: Fit in with an account of the relation between meaning and mind/brain. Why should one expect (or at least hope for) a *semantic* theory to fit into an account of the relation between meaning and mind or brain? Because it would be surprising if the nature of meaning (what meaning *is*) were utterly irrelevant to explaining what it is to grasp or understand meanings, and how grasping meanings can have physical effects. At least, one can imagine differences between x and y that make for a difference between what it is to grasp x and y. For example, understanding x may require skills or recognitional abilities, whereas understanding y may require only propositional knowledge.

I said "mind *or* brain," but in fact I will focus on the brain. And in discussing this matter, I will simply adopt a form of materialism (the "token" identity thesis—that each particular mental occurrence is a physical occurrence).

What is supposed to be in need of explanation about the relation of meaning to the brain? Well, one obvious question is: what is it for the brain to grasp meanings, and how is it that the brain's grasp of meanings has effects on the world? Meanings are (at least apparently) nonphysical abstract objects. And the relation between a brain and the meanings it grasps does not seem to be like the relation between a metal bar and the number of degrees Celsius that is its temperature—a case in which there are proposals about how a change in the value of the temperature can cause, say, expansion of the bar (see Field 1980). Yet the difference between a brain that grasps a certain meaning and a brain that does not makes for a difference in the causal properties of that brain. A brain that grasps the meaning of 'transmogrify' can win a quiz show for its owner, transporting the two of them to a hotel in the Catskills. We need an account of how such a relation between a brain and a meaning can make a causal difference.

Desideratum 6: Illuminate the relation between autonomous and inherited meaning. If there are representations in the brain, as the representational theory of the mind contends, then there is an obvious distinction to be made between them and other representations—for instance, representations on this page (Searle, 1980a; Haugeland, 1980). The representations on the page must be read or heard to be understood, but not so for the representations in the brain. The representations on the page require for their understanding *translation,* or at least *transliteration* into the language of thought; the representations in the brain (some of them, at any rate) require no such translation or transliteration. Let us say that the representations that require no translation or transliteration have *antonomous* meaning, where as the ones that do require translation or transliteration have *inherited* meaning.

Different views of meaning have quite different consequences for the

issue of what a semantic theory could hope to say about either type of meaning. On Searle's view, for example, the most a semantic theory could say about this matter is to give an account of how inherited meaning (*observer-relative* meaning, in his terminology) is inherited from autonomous meaning (*intrinsic meaning,* in his terminology). Explaining autonomous meaning itself, in his view, is simply outside the scope of semantics. The most we can say by way of giving an account of autonomous meaning, according to Searle, is that it arises from the causal powers of the human brain and would arise from any other object (e.g., a machine) that has "equivalent causal powers."

Despite the panoply of views on this matter, there are a few questions whose interest should be agreed on by all who accept the distinction between autonomous and inherited meaning to begin with. The main questions are: What are autonomous and inherited meaning? What is the relation between autonomous and inherited meaning? For example, are they just two different types of meaning, neither of which is derivative from or reducible to the other?[3]

A related question is how a representation with autonomous meaning can mean the same as a representation with inherited meaning. Many philosophers would disparage such a question because of skepticism about synonomy. But it is not clear that those who accept it are caught in the Quinean quicksand. That depends on whether the notion of meaning used in cognitive science must carry with it commitment to *truths* of meaning, and hence commitment to a priori truth.[4]

Desideratum 7: Explain the connections between knowing, learning, and using an expression, and the expression's meaning. Obviously, there is a close connection between *the meaning of a word,* on the one hand, and *what we know when we know or understand a word* and *what we learn when we learn a word,* on the other hand. Indeed, it is intuitively plausible that these italicized descriptions have the same referent (though it would be a mistake to adhere dogmatically to this pretheoretic intuition).

Further, one who has learned an expression (and therefore knows it) automatically has a capacity to use it correctly; also, evidence of correct usage is evidence for knowing the meaning. A psychologically relevant theory of meaning ought to illuminate the connections between knowing/understanding/learning and usage, on the one hand, and meaning on the other.

Desideratum 8: Explain why different aspects of meaning are relevant in different ways to the determination of reference and to psychological explanation. One can distinguish between two aspects of meaning that are relevant to psychological explanation in quite different ways. One type of case involves indexicals, for example:

(1) I am in danger of being run over.
(2) Ned Block is in danger of being run over.

Consider the difference between the beliefs I would express by uttering (1), as compared with (2). Believing (2) cannot be guaranteed to have the same life-saving effect on my behavior as believing (1), since I may not know I am Ned Block (I may think I am Napoleon).[5] So there is an important difference between (1) and (2) with respect to causation (and therefore causal explanation) of behavior.

This observation is one motivation for a familiar way of thinking about meaning and belief content in which, when you and I have beliefs expressed by our (respective) utterances of (1), we have beliefs with the same content. This is the way of individuating in which two lunatics who say "I am Napoleon" have the *same delusion.* Corresponding to this way of individuating belief content, we have a way of individuating meanings in which the meanings of the two lunatics' sentence tokens are the same. This is the way of individuating meanings of tokens that is geared toward sentence types, and thus seems most natural for linguistics—since it makes the meaning of a sentence a function of the meanings of the words in the sentence (plus syntax). Notice that on this way of individuating, utterances of (1) and (2) by me have *different* meanings and standardly express beliefs with *different* contents. Again, this way of individuating is natural for linguistics, since no reasonable dictionary would give 'I' and 'Ned Block' the same entry.

Nonetheless, (1), said by me, and (2) express the same proposition, according to a familiar way of individuating propositions. In a familiar sense of 'meaning' in which two sentence tokens have the same meaning just in case they express the same proposition, (1), said by me, and (2) have the same meaning. If we individuate contents of beliefs as we individuate the propositions believed, the belief I express by (1) would have the same content as the belief I express by (2). Further, the belief I express by (1) would have different content from the belief you express by (1); similarly, the meaning of my utterance of (1) would be different from your utterance of (1).

Call the former scheme of individuation *narrow* individuation and the latter *wide* individuation (cf. Kaplan's different distinction between character and content). Wide individuation groups token sentences together if they attribute the same properties to the same individuals, whereas narrow individuation groups sentence tokens together if they attribute the same properties using the same descriptions of individuals—irrespective of whether the individuals referred to are the same. In other words, narrow individuation abstracts from the question of (i.e., ignores) whether the same individuals are involved and depends instead on how the individuals are referred to.[6] (Note that the question of how individuals are referred to is quite different from the question of how the referrer thinks of the referent. For example, two uses

of (1) have the same narrow meaning (in my sense of the phrase) even if one user thinks he's Napoleon while the other thinks he's Wittgenstein.)

One can think of narrow and wide individuation as specifying different aspects of meaning, narrow and wide meaning. (I am not saying that narrow and wide meaning are *kinds* of meaning, but only aspects or perhaps only *determinants* of meaning.) Narrow meaning is "in the head," in the sense of this phrase in which it indicates supervenience on physical constitution,[7] and narrow meaning captures the semantic aspect of what is in common to utterances of (e.g.) (1) by different people. Wide meaning, by contrast, depends on what individuals outside the head are referred to, so wide meaning is not "in the head." The type of individuation that gives rise to the concept of narrow meaning also gives rise to a corresponding concept of narrow belief content. Two utterances have the same narrow meaning just in case the beliefs they express have the same narrow content.

Note that despite the misleading terminology, wide meaning does not *include* narrow meaning. Utterances of (1) (by me) and (2) have the same wide meaning but not the same narrow meaning.[8]

Narrow meaning/content and wide meaning/content are relevant to psychological explanation in quite different ways. For one thing, the narrow meaning of a sentence believed is more informative about the mental state of the believer. Thus narrow meaning (and narrow content) is better suited to predicting and explaining what someone decides or does, so long as information about the external world is ignored. Thus, if you and I both have a belief we would express with (1), one can explain and predict our sudden glances at nearby vehicles and our (respective) decisions to leap to the side. Wide meanings are less suited to this type of prediction and explanation, because they "leave out" information about the way one refers to oneself. Since the wide meaning of (1) said by me and (2) are the same, if you are told I believe a sentence with this wide meaning (i.e., the wide meaning common to my [1] and [2], you know that I believe that something—me, as it happens, but you aren't told that I know it's me—is in danger of being run over. Thus, information is omitted, since you aren't told how I conceive of the thing in danger. On the other hand, you do know that I believe that something is in danger, so you do have *some* information about my mental state.

From what I have just said, it would seem that narrow meaning includes everything relevant to psychological explanation that wide meaning does, and more. But wide meaning may be more useful for predicting in one respect: to the extent that there are nomological relations between the world and what people think and do, wide meaning will allow predicting what they think and do without information about how they see things. Suppose, for example, that people tend to avoid wide open spaces, no matter how they describe these spaces to themselves. Then knowing that Fred is choosing whether to go via an open space or a city street, one would be in a position

to predict Fred's choice, even though one does not know whether Fred describes the open space to himself as 'that', or as 'Copley Square'.

Narrow meaning has another kind of theoretical import: it determines a function from expressions and contexts of utterance onto referents and truth values.[9] When you and I utter 'I' in (1), there is something we share, some semantic aspect of the word 'I' that in your context maps your token onto you and in my context maps my token onto me.

Let me guard against some misunderstandings. First, as I already indicated, the narrow meaning of 'I' does not include one's conception of oneself. Second, although I have said that there is a shared semantic aspect of 'I' relevant to explaining behavior and a shared semantic aspect relevant to determining a function from context to referent, I do not suggest that these shared semantic aspects are exactly the same. It is an open question whether they are the same, and hence whether 'narrow meaning', as I am using the term, picks out a single thing. On the theory I will be arguing for, the semantic aspect that determines the function from context to referent (and truth value) turns out to be a *part* of the semantic aspect that plays a part in explaining behavior. Thus the latter semantic aspect does *both* jobs. Hence, I will use 'narrow meaning/content' as uniquely referring to the more inclusive semantic aspect. I do want to note, though, that this way of talking carries a strong theoretical commitment. Finally, the narrow/wide distinction as I have described it so far applies to tokens, not types. However, there is an obvious extension to (nonindexical) types.

I will now pause to say what the considerations raised in this section so far have to do with a semantics for psychology. First, a semantics for psychology should have something to say about what the distinction between narrow and wide meaning comes to and, ideally, should give accounts of what the two aspects of meaning are. Second, the theory ought to say why it is that narrow and wide meanings are distinctively relevant to the explanation and prediction of psychological facts (including behavior). Third, the theory ought to give an account of narrow meaning that explains how it is that it determines a function from the context of utterance to reference and truth value.

I have been talking so far about the meaning of sentences with indexicals, but the points I have been making can be extended to names and, more controversially, to natural kind terms. Consider Teen (of Earth) and her twin on Twin Earth, Teen_{te}. The two are particle-for-particle duplicates who have had exactly the same histories of surface stimulations. In various different versions of the story, we are to imagine various differences in their worlds outside the sphere of what has impinged on them. For now, let us suppose their environments are exactly the same, except, of course, that the individuals on the two worlds are distinct—Teen's hero is Michael Jackson, whereas Teen_{te}'s hero is a distinct but indistinguishable (except spatiotemporally) personage. Teen and Teen_{te} each have the thought they would express with:

(3) Michael Jackson struts.

Once again, we can distinguish between two ways of individuating thought contents, and also the meanings of the sentences thought. On one, the narrow scheme, we can talk of Teen and Teen_{te} as having the same thought, and we can talk of them as uttering sentences with the same meaning. If they would both sincerely say "Michael Jackson has supernatural powers," they share the same delusion. This is narrow meaning and narrow content. Alternatively, we can regard the meanings and thought contents as distinct simply in virtue of the fact that Teen is referring to Michael Jackson and Teen_{te} is referring to $\text{Michael Jackson}_{te}$. This is wide meaning and content.

This illustrates same narrow/different wide meaning and content. The case of same wide/different narrow meaning (the case analogous to [1] and [2] above uttered by the same person) is illustrated by 'Cicero orates' and 'Tully orates'. The principles of individuation in these name cases are the same as in the indexical cases, though their motivation is in one respect weaker because it is controversial whether names even *have* meanings. Also, the nomological connection between names and behavior is not as simple as that between 'I' and behavior.

There are two basic facts on which the narrow/wide distinction is based. One is that how you represent something that you refer to can affect your psychological states and behavior. So if you know that Cicero orates and you don't know that Cicero = Tully, you are not in a position to make use of the fact that Tully orates. The second basic fact is that there is more to semantics than what is "in the head." The contents of the head of a person who asserts (3), together with the fact that Michael Jackson struts, are *not enough to determine whether (3) is true or false,* since the truth value depends as well on who 'Michael Jackson' refers to. Imagine that though Michael Jackson is an excellent strutter, his twin cannot strut; the strutting ascribed to his twin by Twin Earth teenagers is actually done by a stuntman. Then utterances of (3) on Twin Earth differ in truth value from utterances of (3) on Earth, despite no relevant differences between teenage heads on the two planets, and despite it being just as much a fact on Twin Earth as on Earth that Michael Jackson struts. (If this seems mysterious to you, note that in the last sentence, I used 'Michael Jackson' as it is used in my language community—Should I talk someone else's language?—and the language community on Twin Earth uses the same expression to refer to a different person.) *Since the truth value of a sentence is determined by the totality of semantic facts, plus the relevant facts about the world, there is more to the totality of semantic facts about the sentence than is in the speaker's head. The "extra" semantic facts are about what the referring terms in the sentence refer to.*[10] But even though there are semantic differences between Teen's and Teen_{te}'s utterance of and thinking of (3), there are important similarities as well—and this is

the main point of this section—that give rise to notions of aspects of content and meaning (narrow content and meaning) *that are shared by Teen and* $Teen_{te}$ and that explain similarities in their (for example) fantasy life and ticket-buying behavior and that determine the function from their different contexts to their different referents.

As in the idexical case, wide meaning and content are not well suited to explaining change of mental state and behavior. The wide meaning of 'Water is wet' (in English—not Twin English) is the same as that of 'H_2O is wet', despite the potentially different effects of believeing these sentences on mental states and on behavior. Further, as Kripke's Pierre example reveals (Kripke 1979), if one's conception of translation is overly referential (allowing 'London' to translate 'Londres' inside belief contexts), one is faced with situations in which one is forced to ascribe contradictory beliefs that are no fault of the believer.[11] In addition, what is shared by Teen and $Teen_{te}$ also determines that one is referring to Michael Jackson, whereas the other is referring to Michael Jackson's twin. What is shared determines a function from context to reference. Had Teen been raised on Twin Earth, she would have been molecule for molecule the same as she actually is (ignoring quantum indeterminacy), but her token of 'Michael Jackson' would have referred to Michael Jackson's twin.[12]

The reader may wonder why I have gone on about this desideratum (on the narrow/wide distinction) at such length. (And I'm not finished yet!) The version of conceptual role semantics that I will be defending characterizes *narrow* meaning in terms of conceptual role. There is another version (Harman 1982) that has no truck with narrow content or meaning. Harman's conceptual roles involve perceptual and behavioral interactions with what is seen and manipulated, that is, objects in the world, whereas my conceptual roles stop at the skin. (So if you don't like all this narrow this and narrow that, you can still appreciate the previous desiderata as motivating a Harmanian version of conceptual role semantics.) I prefer my version, and I am trying to spell out part of the motivation for it.[13] (I will say more about Harman's alternative shortly.)

Consider Putnam's original Twin Earth story. My doppelgänger (again, a physical duplicate)[14] uses 'water' to refer to XYZ. Suppose, along with Putnam, that XYZ is *not* a type of water. Further, we may add into the story ideas developed by Burge (Burge 1979) that show the differences in how our different language communities use words can determine differences in the meanings of our words, even when they do not result in differences in stimuli impinging on our surfaces. Suppose my twin and I both say to ourselves:

> My pants are on fire. But luckily I am standing in front of a swimming pool filled with water. Water, thank God, puts out fires.

If Burge and Putnam are right (and I am inclined to agree with them), there

are substantial semantic differences between my twin's and my meanings and thought contents because of the differences in physical and social environment. Nonetheless—and here, again, is the crucial idea behind my advocacy of narrow meaning and content—*there is some aspect of meaning in common to what he says and what I say (or at least a common partial determinant of meaning), and this common semantic aspect of what we say provides part of a common explanation of why we both jump into our respective pools.* And if current ideas about the representational theory of mind are right, narrow meaning and content will be usable to state nomological generalizations relating thought, decision, and action.

Further, had my twin grown up in my context, his token of 'water' would refer to H_2O rather than XYZ. Thus, as before, it seems that there is some common semantic aspect of our terms that operates in my case to map my context onto H_2O, and in his case to map his context onto XYZ.

The reader may have noticed my shift to the natural extension I described of the narrow/wide distinction from tokens to types. Since 'Cicero' and 'Tully' are standardly used to refer to the same person, we can regard the sentence types 'Cicero orates' and 'Tully orates' as having the same wide meaning. Likewise for 'water' (as used in English as opposed to Twin English) and 'H_2O'.

Let us say that a propositional attitude or meaning ascription is individualistic if it is supervenient on the physical state of the individual's body, where physical state is specified nonintentionally and independently of physical and social conditions obtaining outside the body.[15] I believe that there is an important individualistic scheme of individuation of beliefs, belief contents, and the meanings of the sentences believed. There is a strong element of individualistic individuation in ordinary thought, but its main home lies in scientific thinking about the mind, especially in contemporary cognitive science. I also agree with Burge and Putnam that there is an important nonindividualistic scheme of individuation in ordinary thought. No incompatability yet.

But Putnam, Burge, and others have also argued against individualistic individuation. Putnam's conclusion (1983) is based on an argument that it is impossible to come up with identity conditions on content or meaning, individualistically considered. I don't have identity conditions to offer, but I am inclined to regard this not as an insurmountable obstacle but as an issue to be dissolved by theory construction. My guess is that a scientific conception of meaning should do away with the crude dichotomy of same/different meaning in favor of a multidimensional gradient of similarity of meaning.[16] After all, substitution of a continuum for a dichotomy is how Bayesian decision theory avoids a host of difficulties—for example, the paradox of the preface—by moving from the crude pigeonholes of *believes/doesn't believe* to degrees of belief.[17]

Burge (1984) is arguing mainly against "pan-individualism," the claim that *all* propositional-attitude individuation in psychology is individualistic. However, I am not advocating this doctrine but only the more limited claim that there is an important strain of individualistic individuation in psychology (and in commonsense discourse). Burge has doubts about this too, but the matter can only be settled by a detailed discussion of psychological practice.

Let me mention only one consideration. Psychology is often concerned with explaining psychological differences. The measure of these differences is *variance*.[18] For example, variance in intelligence and other mental attributes and states is ascribed to differences in genes and environment (and interactions of various sorts between these causal factors). Suppose we fill a tour bus with travelers, half from Twin Earth and half from Earth. The Earthlings believe that water is wet and prefer drinking water to gasoline, whereas the Twin Earthlings do not hve these propositional attitudes (because when they think about what they call 'water', they are not thinking about water—they have no term that refers to water). Suppose that the Earthlings and Twin Earthlings do not differ in relevant ways in genes or in the surface stimulation that has impinged on their bodies over their whole lives. Hence, in this population, differences in propositional attitudes cannot be attributed to environment (in the sense of surface stimulation) and genes (and their interactions): the differences in water attitudes are due to something that has nothing to do with differences in the genes or surface stimulations that have affected these people. An analysis of variance would have to attribute a large component of variance to differences in a factor that does not cause any differences in proteins, synaptic connections, or any other physicochemical feature of the body, as do differences in genes and surface stimulations. This would amount to a kind of action at a distance, and this would clearly go counter to the methodology of attribution of variance. (Note that this point could have been formulated in terms of Burge's point about the social nature of meaning rather than Twin Earth.)

I just argued for individualistic individuation of propositional attitude states—for example, beliefs. But there is a gap between individuating beliefs individualistically and individuating belief *contents* individualistically. One might hold that when you individuate belief individualistically, you still have belief of some strange sort; but that content, individualistically individuated, is like a president who is deposed—no longer a president (cf. Stich 1983). I propose to fill the gap as follows.

Where we have a relation, in certain types of cases we have individualistic properties of the related entities that could be said to ground the relation. If x hits y, y has some sort of consequent change in a bodily surface, perhaps a flattened nose, and x has the property of say, moving his fist forward. Of course, the same individualistic property can underlie many different relational properties, and some relations notoriously don't depend

on individualistic properties—for example, 'to the left of'. When content is *non*individualistically individuated, it is individuated with respect to relations to the world (as in the Twin Earth case) and social practice (as in Burge's arthritis example).[19] There is a nonrelational aspect of propositional attitude content, the aspect "inside the head," that corresponds to content in the way that moving the fist corresponds to hitting. This nonrelational aspect of content is what I am calling narrow content. But is narrow content really content?[20]

I find much hostility among philosophers to the ideas of narrow content and narrow meaning. There are many reasons for this resistance that I accept as points of genuine controversy, and about which I am not at all confident about my position. But the worry just mentioned seems to me misplaced, at least as a criticism of conceptual role semantics. The criticism is that I have wrongly assumed that the aspect of meaning or content that is inside the head is something genuinely *semantic*. Jerry Fodor once accused me of a "fallacy of subtraction," that is, of assuming that if you take meaning or content and *subtract* its relation to the world and its social aspect, what you have left is something semantic.

There *is* such a thing as a fallacy of subtraction, of course. If you subtract the property of being colored from redness, you do not get colorless redness. But the issue with respect to conceptual role semantics is merely verbal. Nothing in my position requires me to regard narrow meaning and narrow content as (respectively) *kinds* of meaning and content. As mentioned earlier, I regard them as aspects of or as *determinants* of meaning and content. All that is required for my position is that what I am calling narrow meaning is a distinct feature of language, a characterization of which has something important to contribute to a total theory of meaning (e.g., as indicated in my desiderata). Similarly for narrow content.

Am I conceding that conceptual role semantics isn't really part of *semantics?* The first thing to be said about this question is that it is of very minor intellectual importance. It is a dispute about the border between disciplines; like so many of these disputes, it can be resolved only by a kind of ordinary language philosophy applied to technical terms like 'semantics' (or, worse, by university administrations). Ordinary language philosophy has its place in analyses of concepts that play a central role in ordinary human thought; but application of these techniques to technical terms, where stipulation is the order of the day, is not very illuminating. Nonetheless, I am as willing to quibble as the next person. The correct application of disciplinary terms depends in large part on developments in the disciplines. Often the pretheoretic ideas about the domain of the discipline are left far behind. If meaning indeed decomposes into two factors, then the study of the nature of these two factors belongs in the domain of semantics, even if one or both of them are quite different from meaning in any ordinary sense of the term.

To appeal to ordinary ideas about meaning to argue for excluding narrow meaning from the domain of semantics is like excluding electrons from the domain of the study of matter on the ground that they aren't "solid" and diffract like light.

Further, the role of narrow meaning in determining the function from context to reference and truth value seems especially deserving of the appellation 'semantic'. (I will argue in discussing Desideratum 1 below that narrow meaning—as specified by conceptual role semantics—does indeed determine this function.)

I will continue to talk, as I have, of narrow meaning and narrow content; but I won't mind if the reader prefers to reformulate, using phrases like 'narrow determinant of meaning'.

CONCEPTUAL ROLE SEMANTICS AND TWO-FACTOR THEORY

Conceptual role semantics is not among the more popular approaches, but it has the distinction of being the only approach (to my knowledge, at any rate) that has the potential to satisfy all these desiderata. The approach I have in mind has been suggested, independently, by both philosophers and cognitive scientists: by the former under the title "conceptual role semantics" and by the latter under the title "procedural semantics." (Oddly, these two groups do not refer to one another.) The doctrine has its roots in positivism and pragmatism and in the Wittgensteinian idea of meaning as use. Among philosophers, its recent revival is due mainly to Harman (following Sellars),[21] and Field.[22] Churchland, Loar, Lycan, McGinn, and Schiffer have also advocated versions of the view.[23] In cognitive science, the chief proponent has been Woods,[24] though Miller's and Johnson-Laird's[25] versions have been of interest. The version I like is a "two-factor theory" something like the one advocated by Field,[26] McGinn, (1982), and Loar (1982). (See also Lycan 1981.)

The idea of the two-factor version is that there are two components to meaning, a conceptual role component that is entirely "in the head" (this is narrow meaning)[27] and an external component that has to do with the relations between the representations in the head (with their internal conceptual roles) and the referents and/or truth conditions of these representations in the world. This two-factor approach derives from Putnam's argument (1975, 1979) that meaning could not both be "in the head" and also determine reference. It also takes heart from the Perry-Kaplan points about indexicals mentioned earlier (character and content are two "factors"). The two-factor approach can be regarded as making a conjunctive claim for each sentence: what its conceptual role is, and what its (say) truth conditions are.[28] I will refer to the *two-factor version* of conceptual role semantics as CRS, though

perhaps it should be TFCRS to remind the reader of the two-factor nature of the theory.

For present purposes, the exact nature of the external factor does not matter. Those who are so inclined could suppose it to be elucidated by a causal theory of reference or by a theory of truth conditions. The internal factor, conceptual role, is a matter of the causal role of the expression in reasoning and deliberation and, in general, in the way the expression combines and interacts with other expressions so as to mediate between sensory inputs and behavioral outputs. A crucial component of a sentence's conceptual role is a matter of how it participates in inductive and deductive inferences. A word's conceptual role is a matter of its contribution to the role of sentences.[29]

For example, consider what would be involved for a symbol in the internal representational system, '→', to express the material conditional. The '→' in 'FELIX IS A CAT → FELIX IS AN ANIMAL'[30] expresses the material conditional if, for example, when the just quoted sentence interacts appropriately with:

> 'FELIX IS A CAT', the result is a tendency to inscribe 'FELIX IS AN ANIMAL' (other things equal, of course).
>
> 'FELIX IS NOT AN ANIMAL', the result is a tendency to prevent the inscription of 'FELIX IS A CAT', and a tendency to inscribe 'FELIX IS NOT A CAT'.
>
> 'IS FELIX AN ANIMAL?', the result is a tendency to initiate a search for 'FELIX IS A CAT'.

Conceptual role is *total causal role,* abstractly described. Consider, by way of analogy, the causal role of herring. They affect what they eat, what eats them, what sees them and runs away, and, of course, they causally interact with one another. Now abstract away from the total causal role of herring to their culinary role, by which I mean the causal relations involving them that have an effect on or are affected by human dining. Presumably, some of what affects herring and what they affect will not be part of their culinary role: for example, perhaps herring occasionally amuse penguins, and this activity has no culinary causes or effects. Similarly, elements of language have a total causal role, including, say, the effect of newsprint on whatever people wrap in it. Conceptual role abstracts away from all causal relations except the ones that mediate inferences, inductive or deductive, decision making, and the like.

A crucial question for CRS (*the* crucial question) is what counts as identity and difference of conceptual role. Clearly, there are many differences in reasoning that we do not want to count as relevant to meaning. For example, if you take longer than I do in reasoning from x to y, we do not necessarily want to see this as revealing a difference between your meanings

of *x* and/or *y* and mine. Our reasoning processes may be the same in all inferentially important respects.

Further, CRS must face the familiar "collateral information" problem. Suppose you are prepared to infer from 'TIGER' to 'DANGEROUS', whereas I am not. Do our 'TIGER's have the same conceptual role or not? More significantly, what if we differ in inferring from 'TIGER' to 'ANIMAL'? Does the first difference differ in kind from the second?

CRS has less room to maneuver here than, say, Katzian semantics, since CRS cannot make use of an analytic/synthetic distinction. The problem is that if we make the inferences that define 'cat' just the putatively analytic ones (excluding, for example, the inference from 'cat' to 'is probably capable of purring'), we get a meaning for 'cat' that is the same as for 'dog'. (One could try to distinguish them by making use of the difference between the words themselves [e.g., the fact that 'is a cat' entails 'is not a dog'], but that would at best allow intrapersonal synonomy, not interpersonal synonomy. See Field 1978.) This is not a problem *within* Katzian semantics because Katzians appeal to primitive (undefined) elements of language in terms of which other elements are defined. (See Katz 1972.) The Katzian picture is that you can distinguish the meaning of 'dog' from 'cat' by appealing to the analytic truths that cats are feline (and not canine) and dogs are canine (and not feline), where 'feline' and 'canine' are primitive terms. This move is not available for CRS, since it has no truck with primitive terms: conceptual role is supposed to completely determine narrow meaning. (One qualification: it *is* possible to take conceptual role as a *part* of a theory of the narrow meaning of *part* of the language—the nonprimitive part—while appealing to some other conception of meaning of primitives; procedural semanticists sometimes sound as if they want to take *phenomenal* terms as primitives whose meaning is given by their "sensory content," while taking other terms as getting their meanings via their computational relations to one another and to the phenomenal terms as well [perhaps they see the phenomenal terms as "grounding" the functional structures]. It should be clear that this is a "mixed" conceptual role/phenomenalist theory and not a pure conceptual role theory.)

Without an analytic/synthetic distinction, we would, as I mentioned earlier, have to move to a scientific conception of meaning that does away with the crude dichotomy of same/different meaning in favor of a multidimensional gradient of similarity of meaning (hoping for results as good as those achieved by decision theory in moving from an all-or-nothing notion of belief to a graded notion).

If CRS is to be developed to the point where it can be evaluated seriously, definite proposals for individuating conceptual roles must be framed and investigated. One of the purposes of this paper is to try to make it plausible that CRS is worth pursuing.

What about the social dimension of meaning demonstrated in Burge (1979)? Two-factor theory *can* try to capture such phenomena in the referential factor. For example, perhaps the causal chain determining the reference of my use of 'arthritis' is mediated by the activities of people who know more about arthritis than I do. (See Boyd [1979] for an indication of how to knit the social aspect of meaning together with a causal theory of reference.) Alternatively, two-factor theory may have to expand to three-factor theory, allowing a distinct social factor to meaning. Since my mission is to compare the broad outlines of the view I am espousing with alternative points of view, I will not pursue the matter further (though later on I will take up the question of how the conceptual role factor is related to the referential factor).

It should be becoming clear that CRS as I am conceiving of it is so undeveloped as to be more of a framework for a theory than a theory. Why bother thinking about such a sketchy theory? I think that the current status of CRS is reminiscent of the "causal theory of reference." The root idea of causal theories of reference seems clearly relevant to central phenomena of reference, such as how one person can acquire the ability to refer to Napoleon from another person, even without acquiring many beliefs about Napoleon, and even if most of what he believes is false. Detailed versions of causal theories (Devitt 1981) have not commanded widespread agreement; nonetheless, since the only alternative theories of reference (e.g., the description theory) seem hopeless as accounts of the phenomena just mentioned, we are justified in supposing that the central ideas of the causal theory of reference will have to play a part in some way in any successful theory of reference. I intend the desiderata I've discussed to provide a similar rationale for supposing that the central ideas of CRS must somehow fit into our overall semantic picture.

I should mention that (as with the causal theory of reference) a two-factor conceptual role semantics has been set out in one precise version—that of Field (1977). Though Field's account is very suggestive, I will not adopt it, for a number of reasons. For one thing, Field's account is not quite a conceptual role account in the sense in which I have defined it, since his conceptual roles are not quite causal. Field defines conceptual role in terms of conditional probability. Two sentences have the same conceptual role if and only if they have the same conditional probability with respect to every other sentence. Though Field is not explicit about this, he obviously intends some kind of causal account in terms of the causal consequences of new evidence on degrees of belief. Harman (1982) criticizes Field's account on the ground that it does not allow for revision of belief. Harman's argument, apparently, is that Bayesians merely change their degree of belief rather than changing their mind. That is, Bayesians do not treat new evidence as dictating that they should reject claims they formerly accepted (or conversely), but rather that they should move from a .67 degree of belief in a claim to a .52

degree of belief. I don't find Harman's objection very persuasive; what corresponds to change of mind in the Bayesian perspective just *is* change of degree of belief. The Bayesians reject change of mind in favor of change of degree of belief; this is a theoretical disagreement that is not settled by insisting. However, a version of Harman's conclusion seems quite likely right, but for another reason: in seeing change of mind entirely in terms of change in degree of belief via conditionalization (or generalized conditionalization), the Bayesian perspective (like the logical empiricist views that are concerned with justification rather than discovery) cannot model the kind of change of mind that involves the generation of new hypotheses (this point is most convincing with regard to new hypotheses that involve new ideas). Its not that the Bayesian perspective is in any way incompatible with the generation of new hypotheses, but rather that on the Bayesian account of reasoning, new hypotheses must be treated as "given" via some non-Bayesian process, and so the Bayesian account is importantly incomplete. Conceptual role includes the kind of reasoning in which one infers from evidence against one's hypothesis to an obvious variant deploying a revised version of an old idea, and this cannot be captured wholly within a Bayesian framework.

Even ignoring this matter, Field's account highlights a choice that must be made by CRS theorists, one that has had no discussion (as far as I know): namely, should conceptual role be understood in ideal or normative terms, or should it be tied to what people actually do? As Harman (forthcoming) points out (in another context), accounts of reasoning that involve change of degree of belief by conditionalizing on evidence require keeping track of astronomical numbers of conditional probabilities. (Harman calculates that a billion are needed for thirty evidence propositions.) So any Bayesian account would have to be very far removed from actual reasoning. However, if we opt against such idealization, must we stick so close to actual practice as to include in conceptual role well-known fallacious reasoning strategies, such as the gamblers' fallacy?[31]

I prefer not to comment on this matter, in part because I'm not sure what to say and in part because I am trying to stay away from controversies *within* conceptual role semantics, because the points I want to make can be made on the basis of a version of the doctrine that contains very little in the way of details.

Calling the causal roles CRS appeals to 'conceptual' or 'inferential' shouldn't mislead anyone into supposing that the theory's description of them can appeal to their meanings—that would defeat the point of reductionist theories. The project of giving a nonsemantic (and nonintentional) description of these roles is certainly daunting, but the reader would do well to note that it is no more daunting than the programs of various popular philosophical theories. For example, the causal theory of reference, taken as a reductionist proposal (as in Devitt's but not in Kripke's versions) has the

same sort of charge. And, a rather similar charge falls on "traditional" non-representational functionalism (e.g., as in Lewis's or Putnam's versions), where the causal roles of propositional attitude states are to be described in nonintentional and nonsemantic terms.

Representationalists differ in how important they think the role of English expressions are in reasoning, deliberation, and so forth. At one end of the spectrum, we have the view that English is *the* language of thought (for English speakers). Near the other end, we have those who, more influenced by cognitive psychology, have tended to see reasoning in English as the tip of an iceberg whose main mass is computation in an internal language common to speakers of English and Walburi.[32] On the latter view, the narrow meaning of English expressions is derivative from the narrow meanings of expressions in the internal language. (The dependency would, however, be the other way around for the referential component of meaning, since it is English expressions that are more directly related to the world.) I will not be concerned with this and a number of other disputes that can occur *within* the framework of conceptual role semantics.

In what follows, I shall be quite relaxed about this issue of the role of English in thinking. Sometimes, I will take English to be the language of thought. However, when it is convenient, I will assume that English is used only for communication and that *all* thought is in a language that does not overlap with English, mentalese. When on this latter tack, I will also assume that mechanisms of language production and language understanding establish a *standard association* between English and mentalese expressions. When a speaker formulates a message using 'CAT', language—production mechanisms map 'CAT' onto 'cat'; and when the hearer understands 'cat', the language—understanding mechanisms map it onto 'CAT'.

This standard-association notion can be used to construct a way of individuating conceptual roles in which English expressions have the conceptual roles of the mentalese expressions with which they are standardly associated. Suppose I am told that Felix is a cat and am asked about Felix's weight. I answer "Felix weighs more than .01 grams." I suggest we start with the following simple mechanistic picture. When I hear "Felix is a cat," language-understanding mechanisms produce "FELIX IS A CAT." Reasoning mechanisms produce "FELIX WEIGHS MORE THAN .01 GRAMS," and language-production mechanisms result in the utterance of "Felix weighs more than .01 grams." Now an English sentence and its internal standard associate certainly hve different causal properties. For example, one is visible or audible (normally) without neurophysiological techniques. But we can individuate conceptual roles so as to give them the *same* conceptual roles, simply by (1) taking the relevant causal properties of English expressions as the ones that are mediated by their causal interactions with their standard associates and (2) abstracting away from the mechanisms that effect the

standard association. Then any cause or effect of 'cat' will, for purposes of individuation of conceptual roles, be regarded as the same as a cause or an effect of 'CAT'.

An analogy: Consider a computer in which numbers are entered and displayed in ordinary decimal notation, but in which all computation is actually done in binary notation. The way the computer works is that there are mechanisms that transform the '3 + 4' you enter on the keyboard into an internal expression we can represent as '+ (11,100)'. This is a translation, of course, but we can talk about it without describing it as such, by describing it in terms of the mechanism that computes the function. Internal computational mechanisms operate on this expression, yielding another expression, '111', which is transformed by the translation mechanisms into a '7' displayed on the screen. Now the process by which '3 + 4' yields '7' is exactly the same as the process by which '+ (11,100)' yields '111', except for the two translation steps. So if we (1) ignore causes and effects of decimal digits other than those mediated by their interactions with binary digits in the innards of the machine and (2) abstract away from the translation steps, we can regard the decimal and corresponding binary expressions as having the same computational roles.

Thus, one can speak of the conceptual roles of English expressions, even when adopting the view that internal computation is entirely in mentalese. This will seem strange if your picture of English tokens is inert expressions in dusty books, as compared with the dynamic properties of the internal representations in which all thought is actually conducted. So remember that I am adverting to what the English expressions do when seen or heard.

Let me try to clarify what I am trying to do with the notion of standard association by mentioning some caveats.

(1) The English language is of course a social object. In speaking of the conceptual roles of English expressions, I do not intend a theory of that social object. Conceptual role, you will recall, is meant to capture narrow meaning. Indeed, since causal roles differ from person to person, CRS deals with *idiolect* narrow meaning rather than public language narrow meaning.

(2) The existence of the mechanisms that effect the standard association is an empirical question (though, as Stich [1983, p. 80] argues, something like this idea seems to be part of commonsense psychology). I appeal to empirical work on the "language module"—see Fodor (1983b). Were the empirical assumption to turn out false, a conceptual role theory of (the narrow meaning of) external language could still be given (in terms of the causal interactions between external and internal language), but what would be lost would be the plausibility of a conceptual role theory in which for almost any external expression, one could expect an internal expression with the same narrow meaning. So as to have my empirical eggs in one basket, let me include the assumption of a language module under the rubric of "representationalism."

(3) In order for the notion of standard association to be usable to define conceptual roles, it must be characterizable nonsemantically and nonintentionally. But doesn't this idea founder on obvious facts about the devious road from thought to language, for example, that people lie? The point of my appeal to the language module is that it works (once engaged) without the intervention of any intentional states. Of course, it is used by us in a variety of ways, since we have many purposes in using language. The language module works the same in lying and truth telling; the difference is to be found in the mentalese message. Perhaps confusion would be avoided if one focuses on the use of language, not in communication, but in thinking out loud or in internal soliloquies.

(4) Language production may have to bear more of the burden in characterizing standard association than language perception, since the latter encounters complications with indexicals and the like. When one hears "I'm sick," one doesn't represent it the way one would represent one's own first person thought.

(5) Despite the convention I've adopted of writing mentalese as English in capitals, nothing in the CRS position requires that a sentence spoken have the same meaning as that sentence thought. One can make sense of the idea that in speech one uses the English word 'chase' to mean what one means in thought by the English word 'CHAIR'. Imagine yourself moving to a place where they speak a dialect of English that differs from yours in exchanging the meanings of these two words. If you continue to think in your old dialect but talk in the new one, you would be in the described situation. Consider two quite different scenarios. In one, the new situation never effects a change in your language production/perception module. In communicating, you consciously adjust your words, but in thinking out loud, you talk as before. In the other scenario, the module changes so as to adjust to the external shift. In the former case, standard association will be normal. In the latter, 'chair' will be standardly associated with 'CHASE', and the conceptual role of 'chair' will derive from 'CHASE'-thoughts (involving trying to catch rather than sitting). 'Chair' will have the same conceptual role as 'CHASE'. Neither scenario provides any problem for the view of conceptual role of external language that I sketched. Schiffer and Loar have emphasized that if there is an internal language, a sentence spoken need not have the same meaning as the same sentence thought, but they have been led to conclude that if a language of thought hypothesis is true, it is reasonable to deploy two quite different types of theories of meaning—one for internal language, one for external language. Their concern with external language is with meaning in public language, whereas mine is with narrow meaning in idiolect, so there is no direct conflict. Still, I want to emphasize that a conclusion analogous to theirs for idiolect narrow meaning is mistaken. (See Loar 1981; Schiffer 1981.) This matter will come up again in the section below on what makes meaningful expressions meaningful.

One final point of clarification: Though I am advocating CRS, I am far from a true believer. My position is that CRS can do enough for us (as indicated by the desiderata it satisfies) to motivate working it out in detail and searching for solutions to its problems.

Perhaps this is the place to mention why I am willing to advocate a version of functionalism despite my arguments against functionalism in Block (1978). First, I am impressed by the questions this particular version of functionalism (apparently) can answer. Second, I am now willing to settle for (and I think there is some value in) a theory that is chauvinist in the sense that it does not characterize meaning or intentionality in general, but only *human* meaning or intentionality. Third, the arguments I gave for the conclusion that functionalism is liberal (in the sense that it overascribes mental properties, e.g., to groups of people organized appropriately) were strongest against functionalist theories of *experiential* mental states. I am now inclined to regard intentional mental states as a natural kind for which a functionalist theory may be OK, even though it is not acceptable for experiential states. Indeed, if the domain of CRS is a natural kind, then so is the domain of intentional mental phenomena.

Ironically, this concession to functionalism may make my position harder to defend against thoroughgoing functionalists, since it may commit me to the possibility of intentionality—even intentional states with the same sort of intentional content as ours—without experience. Perhaps I would be committed to the possibility of "zombies," whose beliefs are the same as ours (including beliefs to the effect that they are in pain), but who have no real pains (only "ersatz" pains that are functionally like pain but lack qualitative content). Then I would have to confront the arguments against this possibility in Shoemaker (1984, chaps. 9 and 14). (On my view, pain, for example, is actually a composite state consisting of a nonfunctional qualitative state together with a functional state. Since the qualitative state can be neurophysiologically—but not functionally—characterized, I regard the full account of the mental as part functional, part physiological.) Finally, I believe many of the other arguments that have been advanced against functionalism in its various forms to be defective (see my argument below against Searle).

Two Factors or One Factor?

The version of CRS I have been talking about is a "two-factor" version, in which the conceptual role factor is meant to capture the aspect (or determinant) of meaning "inside the head," whereas the other is meant to capture the referential and social dimensions of meaning.

As I mentioned earlier, Gilbert Harman has been advocating a different version of conceptual role semantics. Harman's version makes do with *one* factor, namely, conceptual role. How does he do without the referential and

social factors? By making his one factor reach out into the world of referents and into the practices of the linguistic community. I have been talking about conceptual roles along lines common in functionalist writing in philosophy of mind. These conceptual roles stop roughly at the skin. Outputs are conceived of in terms of bodily movements or, according to the more scientifically minded; in terms of outputs of, say, the motor cortex (allowing for thoughts in disembodied brains). Inputs are conceived of in terms of the proximal stimuli or in terms of outputs of sensory transducers. By contrast, here is Harman on the subject.

> Conceptual role semantics does not involve a "solipsistic" theory of the content of thoughts. There is no suggestion that content depends only on functional relations among thoughts and concepts, such as the role a particular concept plays in inference. (Field, 1977, misses this point.) Also relevant are functional relations to the external world in connection with perception, on the one hand, and action on the other. What makes something the concept red is in part the way in which the concept is involved in the perception of red objects in the external world. What makes something the concept of danger is in part the way in which the concept is involved in thoughts that affect action in certain ways.[33]

One might speak of Harman's conceptual roles as "long-armed," as opposed to the "short-armed" conceptual roles of the two-factor theorist.

My objection to Harman, in brief, is that I don't see how he can handle the phenomena one would ordinarily think of as being in the purview of a theory of reference without extending his account to the point where it is equivalent to the two-factor account.

The point emerges as one looks at Harman's responses to problems that are dealt with by familiar theories of reference. Consider a resident of Earth who travels to Twin Earth in a space ship. He lands in a body of XYZ; but, ignorant of the difference between Twin Earth and Earth, he radios home the message "Surrounded by water." At first glance, one would think that the Harmanian conceptual role of the traveler's word 'water' would at that moment involve a connection to XYZ, since that is what his perception and action is at that moment connected with. Then Harman would be committed to saying the traveler's message is true—in contrast with the Putnamian claim that his message is false because he is not surrounded by water (but rather twin water). Since Harman accepts the Putnam line, he deploys a notion of "normal context" (Harman 1973), the idea being that the traveler's conceptual role for 'water' is to be thought of as involving the substance he normally refers to using that word.

Another case Harman discusses is Putnam's elm/beech case. (You will recall that the question is how I can use 'elm' to refer to elms when what I

know about elms is exactly the same as what I know about beeches (except for the names). Harman's solution is to include in *my* conceptual role for 'elm' its role in the minds of experts who actually know the difference.

It begins to look as if Harman is building into his long-arm conceptual roles devices that have usually been placed in the theory of reference. The point can be strengthened by a look at other phenomena that have concerned theories of reference, such as borrowed reference to things that do not now exist but did exist in the past. I can refer to Aristotle on the basis of overhearing your conversation about him, even if most of what I believe about Aristotle is false, because I misunderstood what you said. Will Harman deal with this by making his conceptual roles reach from one person to another, into the past, that is, making a causal relation between Aristotle and me—mediated by you, and your source of the word, and your source's source, etc.—part of the conceptual role of my use of 'Aristotle'? If not, how can Harman handle borrowed reference? If so, Harman certainly owes us a reason for thinking that the outside-the-body part of his long-arm conceptual roles differs from the referential factor of two-factor theory.[34] The burden of proof on Harman is especially pressing, given that it appears that one could easily transform a theory of the sort he advocates into a theory of the sort I have been advocating. If you take Harman's long-arm conceptual roles and "chop off" the portion of these roles outside the skin, you are left with my short-arm conceptual roles. If the outside-the-body part that is chopped off amounts to some familiar sort of theory of reference, then the difference between Harman's one-factor theory and two-factor theory is merely verbal.

Conceptual role semantics is often treated with derision because of failure to appreciate the option of a two-factor version, a failure that is as common among the proponents of the view as the opponents. Consider Fodor's critique (1978) of Johnson-Laird's version of conceptual role semantics. Johnson-Laird's version tended in his original article towards verificationism; that is, the roles of words he focused on were their roles in one specific kind of reasoning, namely verifying. Fodor correctly criticizes this verificationism.[35] But I want to focus on a different matter. Fodor objected that the meaning of 'Napoleon won the Battle of Waterloo' could not possibly consist in any sort of a set of procedures for manipulating internal symbols. That idea, he argued, embodies a use/mention fallacy.

> Suppose somebody said: 'Breakthrough! The semantic interpretation of "Did Napoleon win at Waterloo?" is: *find out whether the sentence "Napoleon won at Waterloo?" occurs in the volume with Dewey decimal number XXX, XXX in the 42nd St. Branch of the New York City Public Library'*. . . . " 'But', giggled Granny, 'if that was what 'Did Napoleon win at Waterloo?' meant, it wouldn't even be a question aobut *Napoleon'*. 'Aw, shucks', replied Tom Swift."[36]

Fodor's objection is that if meaning is identified with the causal interactions of elements of language, sentences would be about *language,* not the world.

My defence of Johnson-Laird should be obvious by now. Take the procedures that manipulate 'Napoleon', etc. (or, better, the whole conceptual roles of these words) as specifying *narrow* meaning. Fodor's argument would only be damaging to a theory that took conceptual role to specify what language is *about.* But if conceptual role specifies only narrow meaning, not reference or truth conditions, then Fodor's criticism misses the mark. Were Johnson-Laird to adopt a two-factor theory of the sort I have been advocating, he could answer Fodor by pointing out that the job of saying what language is about is to be handled by the referential component of the theory, not the narrow-meaning component.

A similar point applies to Dretske's rather colorful criticism of remarks by Churchland and Churchland, (1983).

> It sounds like magic: signifying something by multiplying sound and fury. Unless you put cream in you won't get ice cream out no matter how fast you turn the crank or how sophisticated the "processing." The cream, in the case of a cognitive system, is the *representational* role of those elements over which computations are performed. And the representational role of a structure is, I submit, a matter of how the elements of the system are related, not to one another, but to the external situations they "express."[37]

But the cream, according to two-factor theory, is conceptual role *together with* Dretske's representational role. Since CRS puts in Dretske's cream, plus *more,* there is no mystery about how you get ice cream out of it.

The same sort of point applies against criticisms of CRS that take the conceptual role component to task for not providing a *full* theory of meaning. Our judgments of sameness of meaning are controlled by a complex mix of conceptual role and referential (and perhaps other) considerations.[38]

Fodor (1985) points out that the concept of water can be shared by me and Blind Me. He says this presents problems for theories like CRS. He goes on to say:

> The obvious reply is that the properties of causal relations that make for sameness and difference of functional roles are very abstract indeed. Well, maybe; but there is an alternative proposal that seems a lot less strained. Namely that if Blind Me can share my concept of water, that's not because we both have mental representations with abstractly identical causal roles; rather, it's because we both have mental representations that are appropriately connected (causally, say) to *water.*[39]

But the two replies he gives aren't *incompatible alternatives;* CRS can adopt

them both—though I think Fodor is right that the fact that the reference by me and Blind Me is to the same stuff is probably the main thing here. The point is that one cannot criticize a two-factor theory for not doing it all with one factor.

OVERVIEW

The rest of the paper is mainly concerned with showing how CRS satisfies the desiderata and with comparing CRS with other semantic theories in this regard. I will be talking about two quite different (but compatible) kinds of semantic theories: reductionist and nonreductionist. A reductionist semantic theory is one that characterizes the semantic in nonsemantic terms. A nonreductionist semantic theory is not one that is *anti*reductionist, but only one that does not have reductionist aims. These theories are mainly concerned with issues about constructions in particular languages, for example, why 'The temperature is rising' and 'The temperature is 70°' do not entail '70° is rising'. The nonreductionist theories I will mention are possible-worlds semantics, the model-theoretic aspect of situation semantics, Davidsonian semantics, and Katzian semantics. The reductionist theories are CRS; Gricean theories, by which I mean theories that explain the semantic in terms of the mental; and what I call "indicator" theories, those whose metaphor for the semantic is the relation between a thermometer and the temperature it indicates, or the relation between the number of rings on the stump and the age of the tree when cut down. These theories regard the nomological relation between the indicator and what it indicates as the prime semantic relation. In this camp I include views of Dretske, Stampe, Fodor, and one aspect of Barwise and Perry's position.

The reductionist/nonreductionist distinction as I have drawn it does not do justice to Davidson's views. The problem is *not* that Davidson's work on, for example, the logical form of action sentences makes him a nonreductionist, whereas his view about what meaning is makes him a reductionist. As I pointed out, the reductionist and nonreductionist enterprises are compatible, and there is nothing at all odd about one person contributing to both. The problem, rather, is that Davidson has views about what meaning is, thereby making it seem (misleadingly) that he is a reductionist, however, his views of what meaning is are clearly *not* reductionist. (See Davidson [1984, p. xiv], where he describes his project as explaining meaning in terms of truth.) A finer-grained classification would distinguish between (1a) reductionist and (1b) nonreductionist theories about what meaning is and distinguish both types of views of what meaning is from (2) the project of model-theoretic semantics, Davidson's work on action sentences, and the like. In labeling (1a) as reductionist and everything else as nonreductionist, I've unhappily lumped together (1b) and (2), but this is unimportant for my purposes, since I am ignoring (1b) theories.

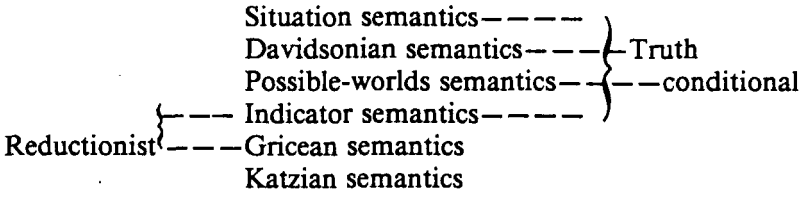
Being reductionist in intent, CRS should not really be regarded as competing with the nonreductionist theories. Nonetheless, I shall be comparing CRS with these nonreductionist theories as regards the desiderata I have listed. To prevent misunderstanding, I want to emphasize that I am not attempting to criticize these nonreductionist theories. Rather, my purpose is to make it clear that they should not be seen as pursuing the same goals as the reductionist theories.

I will also be comparing CRS with the reductionist theories. These theories are in the same ball park as CRS, but most are not genuine competitors. Since CRS in the version I am promoting is a two-factor theory, it requires the partnership of a reductionist truth-conditional theory. Indicator semantics is a candidate. Another candidate that is both truth-conditional and reductionist is Field's interpretation (1972) of Tarski. I won't be discussing it because I know of no claims on its behalf that it is a full theory of meaning—indeed, Field views it as a candidate for the truth-conditional factor of a two-factor theory (see Field 1977). Though I do not regard indicator semantics as a real competitor, I will mention serious problems with the view.

The only circumstance in which the reductionist truth-conditional theories would be genuine competitors to CRS would be if one of them could satisfy a range of desiderata of the sort I've mentioned. I consider it no problem if they can contribute to *some* such desiderata, since there is often more than one way of explaining something. But if some truth-conditional reductionist theory could satisfy *all of them,* the need for the conceptual role component would be brought into question.

The only approach that remains as a genuine competitor is the Gricean approach. I shall not attempt to refute this approach (for one thing, as will appear, it has considerable similarity to mine). I mainly aim to block an argument that anyone who favors a functionalist approach to meaning should adopt some sort of Gricean view rather than CRS.

A brief guide to the semantic theories I will be mentioning: I lump the truth-conditional theories minus indicator semantics plus Katzian semantics together as nonreductionist. Gricean and indicator theories, by contrast, are reductionist.



As you can see, four of the six theories I will be contrasting with CRS are classifiable as truth-conditional. While CRS in the version I am adopting

has a truth-conditional component, it will play little role in satisfying the desiderata. Thus it may seem that I am taking truth-conditional theories to task for not doing something that they were never intended to do. The rationale for the contrasts I will be making is that radical disagreement is so common with regard to matters semantic that there is little consensus about which semantic theories have which purposes. For each of the truth-conditional theories I will mention, claims have been made on its behalf in the direction of satisfying desiderata of the sort I've listed.

Representationalism

Before I go on to discuss how CRS satisfies the desiderata, I want to make sure my representationalism is not misunderstood. I am committed to complex reasoning being a process that involves the manipulation of symbolic structures. I am not committed to the idea that these symbolic structures are *independent* of representational states of mind, mental objects that are viewed by an inner eye. It is convenient to talk in terms of internal representations as if they were literally sentences in the brain (and I do talk this way), but this talk is, of course, metaphorical. My commitment will be satisfied if the representational states themselves constitute a combinatorial system; that is, if they are structured in a way that allows parts corresponding to words to be combined so as to constitute representational states corresponding to sentences.[40]

I am not committed to the manipulation of symbol structures being involved in *all* reasoning, since I want to allow for "primitive" reasonings out of which complex reasonings are built. (E.g., in some computers, multiplication is a symbolic process in that a multiplication problem is "decomposed" into a series of addition problems; but addition itself is not "decomposed" into another type of problem, but rather accomplished by a hardware device, a primitive processor, that contains no internal representations. If you ask how the computer multiplies, you get a representational answer; if you ask how it adds, you do not.) I am not committed to rules for reasoning being themselves represented. Such an assumption involves notorious paradoxes, and in computers we have examples of symbol manipulators many of whose symbol-manipulating "rules" are implicit in the way the hardware works (See Block 1983). I am not committed to any detailed thesis as to what the internal computations are like. For example, I am not committed to any such idea as that in computing '99 + 99 = 198' there is any internal analog of carrying a '1', or any such symbol manipulation of the sort a person might carry out in doing such a sum.

Further, the claim that we are symbol manipulators is intended as empirical and contingent. I find the idea perfectly intelligible and possible that we are "analog" computers whose internal activities involve no symbol manipulation at all. I make the representationalist assumption for two rea-

sons: the most promising line of research in cognitive science is massively committed to representationalism, and it seems to be paying off; and I believe that there are an astronomical number of thoughts that people are capable of having. I would argue that the number of thinkable sentences thirty words long is greater than the number of particles in the universe. Consider the set of entertainable sentences of the following form: $n \times m = q$, where n and m are in the hundreds of billions range familiar from the national budget (twelve figures), and q is twice as long. Many of these sentences are not believable (e.g., nine hundred billion times itself = 0), but each is certainly thinkable. The number of distinct entertainable propositions of the form mentioned is on the order of forty-six digits long. An instructive comparison: the number of seconds since the beginning of time is only about eighteen digits long. I don't see what the mechanism could be by which a person can think any one of such a vast variety of thoughts without some sort of combinatorial system being involved. My representationalist assumption is in the spirit of Smart's claim that pain is a brain state: an empirically based thesis about what reasoning most likely is.

Of the semantic theories I will be contrasting with CRS, only Fodor's version of indicator semantics has a comparable representationalist assumption; nonetheless, I do not think that my representationalism ought to be seen as the key difference between the theory I am advocating and most of the other theories. For one thing, a denotational theory like Fodor's could be framed in terms of assent to English sentences instead of computational relations to internal sentences. Fodor is a sententialist in that he believes that propositional attitude states are relations to internal sentences. But the internal sentences have no privileged *semantic* role in his account. Also, there are nonrepresentationalist avenues towards the type of functionalist-based semantics I am advocating—for example, Loar's and Schiffer's version of the Gricean program. If CRS in the form in which I am advocating it were to meet serious empirical problems because of its representationalism, I would pursue a nonrepresentationalist version.

Question: If my basic commitment is to a functionalist theory of meaning, why don't I *now* adopt a nonrepresentationalist version of functionalism (e.g., the Loar-Schiffer program) instead of pursuing a program based on a risky empirical assumption (representationalism)? Answer: As I shall point out later, even if the Loar-Schiffer program works for natural language, if there is a language of thought not identical to natural language, their theory won't work for *it*. So *both* theories are subject to empirical risk. Theirs is inadequate if representationalism is true, whereas mine is wrong if representationalism is false.

SATISFYING THE DESIDERATA

In the rest of the paper, I shall be mainly concerned with showing how CRS satisfies the desiderata I sketched above and contrasting CRS's treatment with treatments possible for other approaches.

What Is the Relation between Meaning and Reference/Truth?

From the CRS perspective, what this question comes to is: what is the relation between the two factors? Are the two factors independent? Do they fit together in a coherent way?

I think the conceptual role factor is *primary* in that it determines the nature of the referential factor, but not vice versa. Suppose, for illustration, that one of the familiar versions of the causal theory of reference is true. What makes it true? Facts about how our language works—specifically, how it applies to counterfactual circumstances. Kripke convinces us that it is possible that Moses did not do any of the things the Bible said he did, but rather was an itinerant Egyptian fig merchant who spread stories about how he was found in the bulrushes, saw the burning bush, and so on. Kripke is convincing because we use names such as 'Moses' to refer to the person who bears the right causal relations to our uses of the name, even if he does not fit the descriptions we associate with the name. This is a fact about the conceptual role of names, one that can be ascertained in the armchair, just by thinking about intuitions about counterfactual circumstances.

Of course, our names could have functioned differently; for example, they could have functioned as the competing "cluster of descriptions" theory dictates. If that had been how names functioned, it too could have been ascertained by thinking of the right thought experiments, since it would be a fact purely about the internal conceptual role of names. For example, if 'Moses' functioned according to the cluster of descriptions theory, the intuition about Kripke's story dictated by the way names function would be "Oh, in that case Moses doesn't exist—there never was a Moses." What makes the cluster theory wrong is that that just isn't the intuition dictated by the function of our terms—the intuition, rather, is given by: "In that case, Moses wouldn't have done the things the Bible ascribes to him."

(Note that one cannot *identify* the intuition dictated by the function of names with the intuitions we actually have about cases, since there are all sorts of other factors that influence those intuitions. In the early days of the mind-body identity theory, many philosophers voiced the intuition that there was something semantically wrong with "I just drank a glass of H_2O." Presumably, they were influenced by the "oddity" of mixing scientific terms with mundane terms. Using intuitions to isolate facts about the function of names is not a simple matter.)

In short, what theory of reference is true is a fact about how referring

terms function in our thought processes. This is an aspect of conceptual role. So it is the conceptual role of referring expressions that determines what theory of reference is true. Conclusion: the conceptual role factor determines the nature of the referential factor.

Note the crucial difference between saying that the conceptual role factor determines the nature of the referential factor and saying that the conceptual role factor determines reference. I hold the former, but not the latter. The two-factor theory is compatible with a variety of different mappings from a single conceptual role onto aspects of worlds. For example, a word with the conceptual role of our 'water' could map onto one substance here, another on Twin Earth, and another on Triplet Earth. What is in the head—conceptual role—determines the nature of reference without determining reference itself.

If what I've just argued is right, it is easy to see that conceptual role determines the function from context to reference and truth value. It is the referential factor (as described in a theory of reference) that determines that 'water' picks out H_2O on Earth, but XYZ on Twin Earth. For example, on a causal theory of reference, this will be held to be a matter of the causal relation to different liquids in the two contexts. But since the referential factor must take context into account in this way in order to dictate reference, it will determine the function from context to reference.

What Is the Connection between the Meaning of an Expression and Knowing or Learning Its Meaning?

CRS says meaning is conceptual role. If someone uses a word (or a word functions in her brain) that has the conceptual role of 'dog', then the word in question means the same as 'dog'. If a person's brain changes so as to cause a word to be used (by her or her brain) so as to have the conceptual role in question, then she has acquired the concept of a dog (unless she already has it); if the word in question is 'dog' itself or a mentalese standard associate of 'dog', and if the brain change is a case of learning, then she has learned the meaning of 'dog'. Also, CRS allows us to see why evidence for proper use of 'dog' is evidence for knowing the meaning of 'dog'. For a word to have proper use is for it to function in a certain way; hence someone whose word 'dog' functions appropriately thereby knows the meaning of dog; hence evidence of function can be evidence of knowing the word. Finally, CRS allows us to see how knowing meaning is related to our ability to use language. To know the meaning of an English word is for it to function in a certain way, and the obtaining of this function, together with certain psychological facts (e.g., about motivation) explains correct external usage.

The nonreductionist theories should not be regarded as aimed at answering the questions just discussed, but should nonreductionists disagree, they could give a kind of answer (in the metatheory, of course). A theory that

postulates a type of semantic value *V* (e.g., truth conditions, situations, sets of possible worlds, markerese structures) can say that what it is to know or acquire the meaning of a sentence is to know or learn or acquire its *V*. But saying this only shifts the question to what it is to know or acquire *V*'s. Consider the project of producing an account of what it is for 'cat' to acquire its semantic value in the child. If the semantic value is conceptual role, we can at least picture how the project would go. But what would the project be like—if not the same as the one we just pictured—for semantic values like truth conditions, situations, sets of possible worlds, or markerese structures (rather, senses expressed by these structures)? Davidsonians say that to know the meaning of 'Snow is white' is just to know that it is true iff snow is white. But, as Harman has pointed out, saying this just raises the issue of how one represents to oneself that snow is white. If one uses some sort of symbol structure (and how else is one supposed to do it?), the Davidsonian has only pushed the question back a step, for now we want a theory of the meaning of the symbol structure itself.

Further, there is an open question, on these nonreductionist semantic theories, as to how knowing a word's or a sentence's *V* could explain our ability to use the word or sentence appropriately. For example, suppose knowing the meaning of "The balloon burst" is knowing what situation it denotes. But how can knowledge of the denoted situation explain how we use the sentence appropriately?[41] Not that these questions could not be answered by the nonreductionist—for example, they could *adopt CRS*. The point is that the nonreductionist semantic theories I mentioned have no account of their own. (Of course, as I keep saying, this is not a *defect* of these theories.)

Another matter that distinguishes CRS from the nonreductionist theories (and the non-Gricean reductionist theories) is that CRS promises to give a semantic explanation of certain "principles of charity." Many philosophers of language imagine a "radical translation" or "radical interpretation" situation, in which one is trying to interpret utterances (typically, the problem is introduced with an anthropological situation, and then it is observed that the same issues arise in justifying the homophonic translation). As many philosophers have stressed, one must consider one's hypotheses about what the foreign terms mean together with hypotheses about the speakers' beliefs (and other propositional attitudes). It is the "simplicity" of the *total* theory that counts. Now it is often said that it is the *truth* of the alien beliefs that counts (Davidson sometimes says this); but this seems clearly wrong, in the absence of reason to believe that the alien has got things right. A better approach to principles of charity emphasizes coherence. Attribution of irrational belief cannot go on without limit; eventually, one loses one's grip on the content of what one has attributed. But this kind of charity can be explained by CRS. To understand the alien's beliefs, one has to appreciate their inferential roles (or rather, the inferential roles of the symbol structures that express them).

If the mismatch between the alien's inferential roles and our own is too great, there will be no way for us to translate what he says (cf. Loar 1982).

Further, to the extent that inferential role is normative (an issue within CRS, and therefore one I have avoided), there will be rationality constraints on what can sensibly be attributed. These rationality constraints are in no way a by-product of considerations about translation or about a mismatch of conceptual roles; rather, they are a matter of constraints on the conceptual roles that can possibly express concepts.

Let us return to the familiar claim that to know the meaning of a sentence is to know its truth conditions. In any sense in which this claim has substantial content, it is not at all obvious. For example, it is possible to imagine someone knowing the entire set of possible worlds in which a sentence is true without knowing what the sentence means. For the *way the person represents the set of possible worlds* may not capture its meaning. Perhaps it is possible to develop a canonical notation for representing possible worlds. In terms of such a notation, one could develop an ordering of possible worlds, and thus one might be able to exhibit a set of possible worlds via an arithmetical predicate that picks out the right numbers. But if one knows, say, that the prime-numbered possible worlds are the ones in which a sentence is true, does one thereby know its meaning? Further, even if no such ordering exists, one can imagine representing the possible worlds in which a sentence is true in a way that makes use of a motley of devices, different devices for different classes of worlds. Such a representation needn't capture what the worlds have in common in virtue of which they are the ones in which the sentence is true.[42]

Though it is not at all obvious that knowing truth conditions guarantees knowing meaning, the converse claim is more plausible. And, as Harman has pointed out,[43] CRS can explain this in the following way: normal users of language understand certain metalinguistic ideas, such as the disquotational use of 'true', and this is what gives them knowledge of truth conditions. The conceptual roles of 'true' and nonsemantic terms yield knowledge of biconditionals like " 'Snow is white' is true iff snow is white." But even if knowing meaning involves knowing truth conditions, one can hardly jump to the conclusion that knowing meaning *is* knowing truth conditions.

The fertility of the CRS account of learning can be illustrated by its solution to what might be tendentiously called Fodor's Paradox. Fodor's Paradox is posed by the following argument (Fodor 1975):

1. Learning the meaning of a word is a matter of hypothesis formation and testing.
2. When we learn a new English term (e.g., 'chase'), we can do so only by hypothesizing definitions in terms already known (including terms of the language of thought).

3. The history of attempts to define English terms "decompositionally" (e.g., 'try to catch') has been a dismal failure, and there are familiar Quinean considerations that explain why. This suggests that most English terms cannot be so defined.
4. Therefore, when a term like 'chase' is learned, it must be learned by hypothesizing a definition in terms of a *single* term of the language of thought, 'CHASE', which has the same meaning as 'chase'. In other words, the typical word-learning hypothesis has the form: 'chase' means 'CHASE'.
5. Therefore, for most terms of English, we grasp them only because they correspond to (indeed, are standardly associated with) innate terms of mentalese.

I call the argument a paradox because the conclusion is obviously unacceptable; the issue is which premise to give up. Why is the conclusion unacceptable? Could scientific concepts like 'meson' and 'enzyme', as well as technological ideas such as 'monitor', 'zipper', and 'transistor', be *individually* innate? If so, either evolution mysteriously foresaw the concepts needed for science and technology, or else progress in science and technology is possible only with respect to a highly arbitrary, accidentally prefigured vocabulary. Were this the case, one could expect that some accidental modification of some current technological device would produce a new and utterly unintelligible device that we could use the way a two-year-old uses a telephone while confused about whether it is a game in which daddy is somehow hiding inside the phone.

So what premise must go? The first premise is empirically plausible, justified, for example, by appeal to the type of errors children make. Also, hypothesis formation and testing is the only model of learning we have.

Much ink has been shed over the third premise. No doubt readers have made up their minds on the issue, and what I could say in a brief space here would be of no use at all. I shall confine myself to the remark that *if* it has been shown that there aren't many analytic decompositional definitions in natural languages,[44] that doesn't *show* that there aren't many decompositional definitions of natural-language terms in mentalese; but the burden of proof is on those who think mentalese differs from English in this respect.

The premise CRS militates against is 2. According to CRS, the way we learn a new English term needn't be a matter of definition at all. Rather, the CRS picture is that the term (or its newly formed mentalese standard associate) comes to have a certain function. To the extent that hypotheses are involved, they are hypotheses about how the term functions in thought, reasoning, problem solving, and so forth.

One way to see what the CRS proposal comes to is to reflect on how one learned the concepts of elementary physics, or anyway, how I did. When

I took my first physics course, I was confronted with quite a bit of new terminology all at once: 'energy', 'momentum', 'acceleration', 'mass', and the like. As should be no surprise to anyone who noted the failure of positivists to define theoretical terms in observation language, I never learned any definitions of these new terms in terms I already knew. Rather, what I learned was how to *use* the new terminology—I learned certain relations among the new terms themselves (e.g., the relation between force and mass, neither of which can be defined in old terms), some relations between the new terms and old terms, and, most importantly, how to generate the right numbers in answers to questions posed in the new terminology. This is just the sort of story a proponent of CRS should expect.[45]

Note that CRS is not a psychological theory. In particular, though it can tell us that Fodor's second premise *needn't* be true, it is compatible with its actually being true. For it is compatible with the CRS account that the way one learns to use a new term correctly is by linking it to a term one already has that functions appropriately.

Let me now raise a bogeyman that will come up repeatedly: psychologism. Am I just making the verbal maneuver of using 'semantics' to mean the study of the psychology of meaning, rather than the study of meaning proper? As pointed out in connection with the question of whether narrow meaning is genuine meaning, this question is a quibble. However, my answer is that although knowing is a mental state and learning is a mental process, it is not psychologism to suppose that a theory of what meaning *is* ought to be in some way relevant to what it is to know or learn meaning. For example, one can imagine quite different ideas of what good taste is (ranging from a form of knowing how to a form of knowing that) that would engender quite different ideas of what it is to learn good taste. Closer to home, consider the view that philosophy is conceptual analysis contrasted with the view that philosophy is a kind of history (in which heavy emphasis is placed on knowing the texts). These conceptions would lead to different ideas of what it is to learn philosophy.

But how is the idea that meaning is, say, truth conditions, supposed to be in any way relevant to what it is to learn or know meaning? (Unless truth conditions are identified with *verification conditions*, in which case we have a rather unattractive *special case* of CRS in which conceptual role is role in verifying.) The issue of what it is to learn or know truth conditions, or situation denoted, or associated semantic marker, or function from possible worlds to truth values is just as much in need of illumination from a theory like CRS as what it is to learn or know meaning.

I chose the desideratum about learning as the place to bring up the psychologism bogeyman first because this desideratum is perhaps the most psychological of the ones I mentioned; so it is this desideratum for which, if I am just changing the subject, it should be most apparent. My hope is that

exposing the weakness of the psychologism charge here will allow me to pay less attention to it with regard to later desiderata.

What Makes Meaningful Expressions Meaningful?

I will use this section to lay out the basic ideas of the comparisons with the alternative theories, especially the reductionist competitors. So this will be a long section. According to CRS, what makes an expression meaningful is that it has a conceptual role of a certain type, one that we may call "appropriate." The difference between 'cat' and 'glurg' is that 'cat' has an appropriate conceptual role, whereas 'glurg' does not. What gives 'cat' the particular meaning that it has is its particular conceptual role. The difference between meaningful expressions with different meanings ('cat' and 'dog') is a conceptual role difference *within* the category of appropriate conceptual roles.[46]

The dominant perspectives in semantics—possible worlds semantics, situation semantics, and the approaches of Davidson and of Katz, can be used to give responses to my questions that look just as good at first glance. Suppose they say, for example, that what makes a meaningful sentence meaningful is that it has truth conditions, or a set of truth values in possible worlds, or an associated (sense expressed by a) markerese structure, or a denoted situation. But such answers just *put off* the semantic issue. For now we want to know what it is that makes for the difference— what it is in virtue of which there *is* a difference—between sentences that *have* and sentences that *lack* truth conditions, truth values in possible worlds, associated markerese structures, or denoted situations, and to these questions these nonreductionist perspectives have no answers.[47]

Of course, one can also ask of CRS what the difference is between sentences that have and sentences that lack conceptual roles. But CRS has an answer: certain causal properties. And if the questioner wants to know why sentences have the causal properties they do, again there are answers, at least in principle, to be sought, of the same sort that one would give to "Why do genes have the causal properties they have?"

What the difference comes down to is that CRS aims for a reductionist account, indeed, a naturalistic-reductionist account, in proposing to explain a semantic property in terms of a naturalistic, nonsemantic property: causation. CRS's reductionism and naturalism allow it to promise an answer to "What makes a meaningful expression meaningful?" The semantic approaches mentioned in the paragraph before last, being nonreductionist, cannot answer this question.

Although the dominant views in semantics should be regarded (in my view) as just not directed towards the sort of question typified by this desideratum, it should be noted that they often seem to be responding to much the same motivation that lies behind a naturalistic-reductionistic account. For instance, Davidsonians, though not reductionists, make much of the

claim that a theory of the sort they favor will allow a deduction from a finite nonsemantic base of a specification of truth conditions for any indicative sentence.

On the whole, most of the standard approaches have been primarily concerned with the *relations* among meanings, not the nature of meaning itself. For example, the standard approaches have been concerned with an aspect of compositionality: how the meanings of larger elements such as sentences are related to the meanings of smaller elements such as words. Another sort of issue motivating the standard theories is what we can tell about the logical form of "Sam ate with a fork" from the fact that if it is true, so is "Sam ate." Another issue (one that has been a stumbling block for possible-worlds semantics and one that situation semantics hopes to make progress on) is what the relation is between the semantic value of 'Grass is green' and 'John believes that grass is green'. But these questions can be and have been discussed without ever broaching the issue of what it is in virtue of which expressions have their meanings in the first place.

The main aim of most of the standard approaches to semantics has been to *correlate* meanings with certain objects, so that relations among meanings are mirrored by formal relations among the corresponding objects. These approaches have often been concerned with a purely *descriptive* project, a kind of "curve-fitting," not with explaining the nature of meaning.

The major tradition within this conception of semantics is well described in Barwise and Perry (1984).[48]

> We have intuitions about the logical behavior of a certain class of sentences. With attitudes reports, these are largely intuitions about the phenomenon of "opacity": reluctance to substitute co-referential terms and the like. We codify these intuitions in a set of logical principles, and then semantics consists of finding a collection of plausible set-theoretic models that makes the logical principles come out correct. I think this is the traditional conception in semantics, and it is the setting for Montague Grammar, but it is what I would now call the thin conception of semantics.

As suggested earlier, the Barwise and Perry effort to produce a semantics that satisfies a richer, "thicker" conception of semantics can be seen as moving on two fronts: one involves model-theoretic ideas (e.g., the idea of a partial model), the other a kind of indicator semantics (discussed later in this section). Another aspect of the thickness that Barwise and Perry seek is to make semantics compatible with commonsense psychology, for example, to avoid the possible-worlds semantics problem that one would seem to believe everything logically equivalent to what one believes.[49]

Now Barwise and Perry (1984) have advocated a functionalist theory of propositional attitudes. Perhaps they (and some Davidsonians, Katzians,

and possible-worlds semanticists) envision a two-stage process of semantic theorizing: first, a nonreductive account of meanings, and second a reductive account aimed at desiderata something like the ones I have mentioned. Theorists in these traditions have not, however, put forward second-stage theories. I know of only two types of reductionist approaches to semantics other than CRS (and the causal theory of reference, which I am not discussing in any detail); after considering an objection, I shall sketch these approaches and their relation to CRS.

It may be objected that I have confused:

(i) In virtue of what is a particular token/ches/-noise an utterance of the English word 'chase' meaning, of course, 'try to catch')?

(ii) In virtue of what does the English word type 'chase' mean 'try to catch'?[50]

It may be said that (ii) has no nontrivial answer—it is part of what it *is* to be the English word 'chase' to mean what it does.[51] Asking (ii), on this view, is like asking what makes the number two even. If it weren't, it wouldn't be the number two. (Or: 'Two is even' is analytic.)

On the objector's view, the problem I raise does not *disappear,* but is rather transformed. Instead of asking what it is in virtue of which 'chase' means what it does, I must ask what it is in virtue of which a token is of the type *'chase' in English,* with the meaning that that word necessarily (?) has. Since the problem survives, I suppose that the real objection here is that the question I raise (being about a token) is really pragmatic rather than semantic.

Perhaps some perspective can be gained by contrasting the question about language with the question of why, in the American system of government, cabinet officers are approved by the Senate but presidential advisors are not? Here there seems little utility to seeing the American system of government as an abstract object that has this property necessarily (or analytically). It is not helpful to see the question as one about whether a certain token system is a token of the type "the American system of government." But is language more like a political institution or more like mathematics? This question won't get us very far. What issues belong in pragmatics as opposed to semantics is a matter to be settled by finding out which way of dividing up issues makes the most theoretical sense, not by consulting intuitions about whether language is more social than mathematical.

The important point against the objection is that it is a mistake to see the contrast the objector raises as hinging on the type/token distinction. This becomes especially obvious when one is reminded that the English language is in constant flux. 'Yuppie' has no meaning in English-1982, but it has a meaning in English-1985. And 'chase' may mean something different in English-1988 from what it does in English-1985. If a word's meaning is a

necessary—though language-relative—property of the word, then (1) we must regard different dialects and language stages as in the relevant sense, different languages; and (2) we must recognize a sense of 'word type' in which word types are language specific. So we cannot speak of one word's different meanings in different dialects. But this is just a peculiar way of talking. As we have seen, there is a natural use of the notion of a word type in which we *can* speak of one word as having different meanings in different dialects (as is the case with 'yuppie'). So, deploying the notion of word type in the latter way, the question of why the word type 'chase' means 'try to catch' in English-1985 is not trivial.

The Reductionist Alternatives

There are two competing families of approaches to semantics that *are* reductionist,[52] and hence that *do* have genuine answers to the questions posed in the desiderata I've been talking about. One of them is the approach of reducing meaning to *mental content.* Call this type of approach "Gricean." The Gricean approach as developed by Grice himself, and later Schiffer, reduces speaker meaning to the content of speaker's intentions. For the speaker to mean such and such by what he says is for him to intend his utterance to affect the propositional attitudes of hearers in certain ways. Sentence meaning, on this theory, can be reduced to speaker meaning via a conventional correlation between sentences in the language and communicative intentions. This conventional correlation makes it practicable for a speaker to use certain sentences to produce certain effects in hearers.[53]

Searle has an approach that is Gricean in my sense, in which the intention isn't communicative but rather an intention to produce an object with certain "satisfaction conditions" (Searle 1983a). A rather different sort of Gricean approach was taken by Ramsey, who attempted to reduce the meaning of an item of language to the beliefs that would be expressed by that item.

Gricean approaches have been enveloped in controversies, none of which will be discussed here. Nor is this the place for a full-dress comparison between Gricean and conceptual role accounts. However, there are a few points of comparison that can be made rather briefly. Although I do not want to belittle the Gricean accomplishment, without a naturalistic account of the mental, the Gricean approach has little to contribute to the project I am discussing. One who is concerned with the questions I have been asking about meaning will be equally concerned with corresponding questions about intentional content. Consider, for example, the three questions involved in the desideratum currently being discussed:

1. What is it about a meaningful expression that makes it meaningful?
2. What is responsible for an expression's having the particular meaning it has?

3. What is the difference between expressions with different meanings in virtue of which they have different meanings?

The Gricean faces corresponding questions about intentional content, viz.:

1′. What is it about a contentful state that makes it contentful?
2′. What is responsible for a state's having the particular intentional content it has?
3′. What is the difference between states with different intentional contents in virtue of which they have their different contents?

In the light of this problem, Griceans have a number of options. First, they could simply regard intentional content as primitive—in other words, regard questions like 1', 2', and 3' as having no answers. For Griceans to take this line would be to give up on satisfying the desiderata I've been talking about. This is the nonnaturalistic option I mentioned. Another line would be to pursue some nonfunctionalist reductionist strategy, such as physiological reductionism. This is an unpromising tack (see Fodor 1974), and it is especially unattractive if one is interested in a semantics that might apply to the language use of an intelligent computer or computerlike machine, if we ever construct one.[54]

Another option is Searle's reduction of intentionality to *the brain or whatever has "equivalent causal powers."* The wild card of "equivalent causal powers" allows Searle to avoid the usual drawbacks of physiological reduction. For example, the theory is not chauvinist because it allows for the possibility that the control systems of intelligent machines can have causal powers equivalent to ours. However, the other side of the coin is that the theory is far from naturalistic. To say a machine has causal powers equivalent to those of the human brain is only to say that the machine has causal properties that result in intentionality. So Searle must either (1) regard intentionality as primitive, in which case he has not answered the questions I am talking about, or (2) he must give some nonintentional analysis of "equivalent causal powers." It is clear that Searle takes option (1). That is, he has no intention of giving a reductionist theory of intentionality, though he takes physicochemical properties of *each being* that has intentional states to cause that being's intentional states. (See Searle 1984.)

Searle repeatedly *says* that it is an empirical question whether a given machine has equivalent causal powers, but the careful reader discerns that it is an empirical question only in that the machine itself *will know* if it does indeed have equivalent causal powers.[55] The crux of the disagreement between Searle and me is not about whether a sapient and sentient machine will have to have innards with causal powers equivalent to those in us (we agree on this); the crux rather is whether some sort of functionalist thesis is true of us. For if intentionality can be characterized functionally, then *the way to make a machine with intentional states is to make a machine functionally*

equivalent to us—the equivalent causal powers of the machine's brain will come along for free. Searle's argument against functionalism is his "Chinese room" argument, to be discussed briefly later in this paper.

There is one final "methodological" point to be made against Searle. One should not adopt his view without proper exploration of the alternatives, since if Searle's account is true, the sciences of mind and meaning would seem to be severely limited. In particular, it is hard to see how science (or philosophy) could ever tell us anything substantive about what the source of autonomous meaning or intentionality is.

Another Gricean option is that championed by Schiffer and Loar (and perhaps Grice): they couple the reduction of meaning to the mental with a functionalist reduction of the mental. A major difference between the functionalism-based Gricean theory and CRS is that the Gricean theory is not committed to any sort of representationalism, even of the weak sort that CRS is committed to (viz., that thoughts have recombinable ingredients). This difference between the Loar-Schiffer account and the CRS account is a disagreement about the empirical facts about how the mind works (or about how much philosophical ice such empirical facts cut), not about the functional source of meaning.[56] In sum, in one version Gricean theory is not a competitor to CRS; in another version, it is a competitor but has drawbacks; and in another version, it differs with respect to representationalism and, of course, the details of the Gricean reduction in terms of intentions, as well as the focus on public language meaning as opposed to idiolect narrow meaning.

I shall now turn to an argument by Loar (1981, chap. 9; 1983) against the sort of view I am advocating. As I understand it, Loar's argument is that a theory of meaning should not depend on a speculative psychological claim such as representationalism. So Loar advocates the Gricean reduction of external language to mentality (coupled with a functionalist reduction of the mental). If representationalism happens to be true, Loar favors what amounts to a conceptual role semantics theory of the internal language (though not external language). My objection is simple: if representationalism is false, CRS is certainly false. But if representationalism is true, Loar is stuck with an intention-based semantics for external language plus a conceptual role semantics for internal language—whereas CRS makes do with the latter type of semantics for *both* types of language. (Of course, Loar is concerned with public meaning rather than narrow idolect meaning, but this fact does not play any direct role in his argument.) So, if representationalism is true, the Loar-Schiffer account seems at a disadvantage.

Is there some way in which the Gricean account could be extended to internal language? Computation in internal symbol systems appears to be of a rather "automatic" sort which gains efficiency through inflexibility.[57] For example, if one memorizes a list of six letters, say 'UEKNMG', and one is

asked whether 'E' is on the list, one does an "exhaustive" serial search, looking at all six letters, one by one, even if 'E' is the first letter in the list.[58] (This is one of the better tested results in all of cognitive psychology.) Is it at all plausible that one forms an *intention* to look at all the items, or to do an exhaustive serial search? Further, even if the uses of the internal system are intentional in some sense, surely the intentions are not intentions to *communicate,* as in the standard Gricean theories.

But what if the internal symbol system *is* English (that is, the same as whatever external language is spoken)? Can the Gricean then avoid the problem of the last paragraph by giving a theory of meaning for English, and simply postulating that sentences in the language of thought have the same meaning as in English? First, it is not at all obvious that the meaning of English as used in thought (if it is used in thought) is somehow derivative from its use in communication. Why not the other way around? Second, and more importantly, I have talked as if it is perfectly possible that English is the language of thought, but this is simply *not* in the cards. For one thing, external language is radically ambiguous, both syntactically and semantically. If there is no confusion in *thought* as between financial banks and river banks, then one word in the internal system presumably does not carry both meanings. And if someone says "I tire of visiting relatives," knowing full well whether relatives are visiting her or whether she is the visitor, then it is doubtful that the English sentence could be the vehicle of the thought. (But see Block and Bromberger 1980.) From what I've said thus far, one might suppose that the language of thought might be a kind of regimented English (e.g., syntactic trees with English terminal nodes, as suggested for part of the language of thought in Harman, [1970]). But, at most, some sort of regimented English could be *part* of the language of thought. (Indeed, although there is controversy over whether English is part of the language of thought, there is none over whether English is the *whole* of the language of thought.) For example, there is enormous evidence for representations in mental imagery (see Block 1983 for discussion and references to the literature); and it is quite out of the question that these representations are in English (none of the defenders of the view that the representations of imagery are languagelike have suggested such a thing). When one looks in any detail at what a languagelike representation would have to be to play the role of representations of imagery, this is obvious. Nor would any such suggestion be remotely sensible for the representations of early vision (see Marr 1982).

If English is part of the language of thought, it would seem especially peculiar to treat the semantics of that part of the language of thought so differently from the semantics of external English.

In sum, the Griceans cannot claim that their account is to be preferred to CRS on the ground that their account has no empirical vulnerability, since both accounts have an element of empirical vulnerability. Nonetheless, the

choice between the two approaches seems mainly a matter of philosophical metatheory. If one wishes to insulate one's semantics from experimental falsification, while being willing to tolerate ad hoc addition of components to handle experimental discoveries, the Loar-Schiffer perspective is better. If one is interested in a semantics based on the best empirical theories extant, CRS is better.

There is a second family of reductionist approaches to semantics that could be claimed to satisfy my desiderata: what I called "indicator semantics." Dretske (1981) and Stampe (1977) have similar versions, which I believe have been refuted by Fodor (1984), who has his own version of the view (Fodor 1984, forthcoming). Barwise and Perry (1983) have a view that has affinities to that of Dretske and Stampe, which I will not be able to discuss in detail here.

Dretske and Stampe say what it is for a sentence S to have the content that T in terms of tokens of S carrying information about T; carrying information, in turn, is cashed in terms of a nomological relation between S's and T's (roughly, an S nomologically requires a T).[59] Fodor objects that if error is possible, then a non-T can cause a tokening of S; but then why should we regard T as the state of affairs with which S is nomologically correlated when S has a *better* correlation with the disjunctive state of affairs whose disjuncts are T and the non-T state that causes S? So it seems that, on the Dretske/Stampe view, error is not possible.

Barwise uses the type/token distinction to deal with this problem. Suppose Ed "says 'It is 4 p.m.' at 4:30. While we can truly report that *Ed means what he says,* we can also truly report that *Ed's statement does not mean that it is 4 p.m.*"[60] Barwise's claim is that 'means' is ambiguous: there is one sense appropriate to tokens, another to types. A false token does not convey the information (this is the sense of meaning appropriate to tokens) conventionally associated with the corresponding type. (What about false sentence types? According to Barwise and Perry, it is only tokens [e.g., utterances] of sentences that have truth values, not sentence types.)

But I don't see how Barwise and Perry propose to avoid Fodor's objection in giving an informational account of sentence type meaning. One often gets the impression that their theory is that the meaning of a sentence type is the information *normally conveyed* by tokens of it. But what could 'normally' come to here? This cannot be shorthand for information conveyed by *true uses,* since that would ruin any attempt to give an account of the semantic in nonsemantic terms. If 'normal' is some sort of appeal to what is usual, however, Fodor's problem stands in the way. The correlation between tokens of S and the disjunction of T and pseudo-T states of affairs (ones that mislead people into false assertions of S) will inevitably be better than the correlation between S and T itself. Indeed, it is not hard to think of sentences whose assertions are more often false than true (e.g., famous last words). If 'normal'

is some sort of appeal to the conventional, Barwise and Perry owe us an account of how that is supposed to connect with information conveyed and how they expect to avoid an analysis of conventionality in terms of intentional notions (as in Lewis's analysis). If it is a teleological notion, my guess is that their account will succumb to the kind of criticism now to be raised against Fodor's own account. (I've been assuming that Barwise and Perry do aim for an account of meaning in nonsemantic and nonintentional terms. This conception seems to me to permeate Barwise and Perry (1983), though it is never explicitly announced. In recent conversations with Barwise and Perry, I gather that they do not take themselves to be aiming for an account of meaning that is reductionist in this sense.)

Fodor's own view attempts to captitalize on the very fact that torpedoes the Dretske-Stampe approach. The basic idea is that, in a sense, error is not possible. The aim of Fodor's theory is to give a naturalistic account of what it is in virtue of which a sentence has the truth condition it has—what *makes* a sentence have the truth condition that it has. Some examples of theories that are in the same ball park: (1) the British empiricist theory that what gives a mental representation its truth condition is *resemblance* between the representation and the state of affairs, (2) the Skinnerian theory that what makes T the truth condition of S is that T is the discriminative stimulus of S. These are both false doctrines, for well-known reasons, but they are nonetheless naturalistic.

Fodor's task is one that many writers have seen the need for. As Field (1972) pointed out, the Tarskian approach, on one construal, yields the truth conditions of sentences only by means of *lists* of the referents of singular terms and the denotations of predicates. ('Boston swelters' is shown to be true only because the object that is listed as the referent of 'Boston' is in the set that is listed as the denotation of 'swelters'.) However, serious suggestions for solving this problem are thin on the ground. The only remotely plausible views I know of are the indicator semantics approach (common to Fodor, Dretske/Stampe, and Barwise and Perry); and Tarski's approach, construed as in Field (1972), together with a naturalistic theory of reference such as the causal theory. (Field's construal of Tarski is as giving a way of reducing truth to primitive denotation.)

The heart of the theory is an account of the truth conditions for mental sentences.[61] The account makes use of the claim that believing is a computational relation between a person and a mental sentence. (This computational relation is described below as the sentence being in the "belief box.") The claim is that what it is for T to be the truth condition for a mental sentence M is:

(1) If the cognitive system is functioning as it is *supposed* to; and
(2) idealizing away from epistemic limitations, then *M is in the "belief box"* $\longleftrightarrow$ *T*.

There are two "wheels" that drive this account: the teleological wheel, indicated by the 'supposed' in condition (1), and the epistemic idealization wheel. The idea behind Fodor's account is that there are cognitive mechanisms that are designed to put sentences in the belief box if and only if they are true. Error results when these mechanisms fail, or when epistemic conditions are less than ideal. Thus, if one can spell out the teleological notion and say what epistemically ideal conditions are in a naturalistic way, one will have a naturalistic theory of truth conditions.

There are serious problems with each of two "wheels." Let us begin with the epistemic idealization. One sees how it is supposed to go for cases of things that are too small to see, or happened too far away or too long ago. In these cases, what Fodor imagines we idealize away from is how big we are, where we are, or when we are. The idea is that if epistemically ideal conditions held, one's nose *would be rubbed in the truth*; then mechanisms whose function it is to make one see the truth would take over, and one would indeed see the truth.

But what about statements to the effect that space is Riemannian, or that some quarks have charm, or even that one is in the presence of a magnetic field? Here, it is not enough to suppose one's nose is rubbed in the truth, for its no use having your nose rubbed in the facts—you have to come up with the right theory, too, and you have to know that it is the right theory. Imagine that in the long run the evidence converges on a Riemannian geometry for the universe The ideal scientific community will only believe in this claim if someone *thinks* of it. After all, it is quite intuitive to suppose that there is exactly one parallel to a given line at a given point, as Euclidean geometry tells us. No *series of measurements* can guarantee that anyone thinks of (or takes seriously, even if they think of) claiming that the Euclidean parallel postulate is false. To make a long story short, I don't see how such theoretical statements can be handled without in one way or another abandoning naturalism—for example, appealing to some sort of magical machinery or smuggling something semantic into the specification of the epistemic idealization. Suppose that whenever a "theoretical" property of the sort I just mentioned is raised, the Fodorean constructs an idealization in which humans have a perceptual detector that detects this property. Nothing semantic need be smuggled in with the description of these detectors: they say '*p*' if and only if *p*. With such detectors, if your nose is rubbed in a fact, you will perceive it to obtain. But this response abandons naturalism. We have no idea how such detectors would work or even whether they are possible. Appealing to them is like saying: "Aha, what makes *T* the truth condition for *M* is that an omniscient wizard (i.e., one who believed '*p*' if and only if *p*) would believe *M* if and only if *T*." You don't get a naturalistic account of truth conditions by appealing to the imaginary behavior of an imaginary being.

Idealizing often starts with something familiar and envisions a systematic change. So, in the last paragraph, we started with normal perceptual detectors and imagined them getting better and better (or, alternatively, more and more numerous). Another idea is to try envisioning systematic change in our theorizing mechanisms. Of course, we need a nonsemantic characterization of the ideal theorizing mechanisms. It won't do to say they find the *right* theory, since that is a semantic notion. Perhaps we can simply envision mechanisms that construct all possible theories and choose the simplest of them that is compatible with the data. The problems here are complex, and I can only hint at them. I would argue that on any formal notion of simplicity (e.g., one that involves counting symbols), it just is not true that the simplest theory is true. And even if the simplest theory were true, this assumption—which, of course, is a semantic assumption—would be part of the account. So, the account would not be naturalistic.[62] On the other hand, a *semantic* conception of simplicity (e.g., one that involves the concept of truth) won't be naturalistic either.

The second wheel driving Fodor's account is the idea that the cognitive system is *supposed* to function in a certain way. How is this teleological talk supposed to be understood? (Anyone who has read the current literature on teleology knows that promising suggestions are hard to find.)[63] Sometimes Fodor talks in terms of a notion of teleology provided by evolutionary theory. The cognitive system is supposed to function a certain way in that that is what evolution designed it to do.

One problem is that one cannot rely on evolution in such a simple way, since one can imagine a molecule-for-molecule duplicate of a baby who comes into being by chance and grows up in the normal way. Such a person would have language with the normal semantic properties, but no evolutionary "design."

Quite a different type of problem comes in through evolutionary theory itself. I think it is now quite generally accepted among evolutionary biologists that one cannot suppose that every phenotypic (i.e., actual) characteristic of an organism is an optimal design feature (in any nontrivial sense), given the environment.[64] To take a rather extreme case, for purposes of illustration, consider the phenomenon of "meiotic drive." Normally, each of a pair of genes has an equal chance of ending up in an offspring: if you have one blue-eye gene and one brown-eye gene, the chance that your child will get one of these from you is equal to the chance that it will get the other. But there are some known cases of genes—the mouse t-allele, for example—that beat out whatever gene they are paired to, thus propelling themselves into the next generation. Any such gene that does not have lethal effects on the phenotype is likely to spread in a population very quickly, even if it has suboptimal pheonotypic effects. The upshot is that there are known mechanisms (of which this is only one of many examples) that could have the effect

of producing cognitive mechanisms that aim, to some extent, at properties of beliefs other than truth.[65]

One final point about Fodor's account. One peculiar fact about it is that it does not exploit the compositional structure of the language at all. (This is especially odd in view of the fact that Fodor's representationalism gives him objects in the belief box ripe for compositional exploitation.) In this respect, it is markedly inferior to Field's proposal (mentioned above), in which the truth conditions for sentences are built up out of naturalistic analyses of reference and denotation. This feature of Fodor's theory renders it vulnerable to the following problem (I am indebted here to Michael Bratman): If S and S' are nomologically correlated states of affairs, then on Fodor's analysis, any sentence that is mapped onto one of them will be mapped onto the other. Consider, for example, the correlated properties of electrical and thermal conductivity (whose correlation is expressed in the Wiedemann-Franz Law). Let us agree with Fodor that it is the function of the cognitive mechanisms to put 'The electrical conductivity is rising' in the belief box (in ideal epistemological circumstances) iff the electrical conductivity is rising. But since the right-hand side of this biconditional is true iff the thermal conductivity is rising, the left-hand side will be true iff the thermal conductivity is rising. So Fodor's theory will not distinguish between the semantic values of 'The thermal conductivity is rising' and 'The electrical conductivity is rising'. I don't see how such a problem can be dealt with without going to a compositional story (e.g., by adding a conceptual role component to the theory).[66]

Let me summarize. I've mentioned two types of reductionist theories—indicator semantics and Gricean semantics. (I've also mentioned the causal theory of reference, but I haven't compared it with CRS since it is not normally thought of as a full semantic theory. It—like indicator semantics—is a candidate for the referential-truth-conditional factor of a two-factor theory.) I've mentioned and endorsed Fodor's reason for thinking one version of indicator semantics won't work, and I've given some reasons to be dissatisfied with Fodor's theory. I've mentioned a few versions of Gricean theory, arguing that Searle's version isn't naturalistic (and so isn't a competitor to CRS); and I have countered an argument that the Grice-Schiffer-Loar version should be preferred to CRS because the former, unlike the latter, does not depend on what psychologists find out about mental representation.

Before I go on to the next desideratum, I shall very briefly consider an objection to the whole enterprise: I have been comparing a conceptual role theory of *narrow meaning* with theories that have conceptions of meaning that are quite different from narrow meaning (and also from one another's conceptions of meaning). Isn't this comparing apples, oranges, and mangoes?

Reply: (1) It would not change my points were I to switch from talk of narrow meaning to talk of meaning. Since meaning, on my view, is a pair of

factors—the narrow meaning factor and the referential factor—to talk in terms of meaning would be to talk in terms of both factors of the two-factor account of meaning rather than just one of the factors (narrow meaning). After all, I chose the desiderata to exhibit strengths of the conceptual role factor of the two-factor theory, and I will be exercising the two-factor theory's right to introduce the referential factor where relevant. (2) It is true that different semantic theories differ in their conceptions of meaning, but that does not make comparison illegitimate. Vienna and New Delhi differ in their conception of dessert, but that won't stop me from preferring strudel to gulabja.

Why Is Meaning Relative to Representational System?

The CRS explanation of this relativity is simple. The conceptual role of a symbol is a matter of how it *functions* in a representational system (for this reason, conceptual role is sometimes called "functional role"). How a representation functions in a system depends, of course, on the system. If meaning is function, as CRS dictates, then meaning is system relative.

The nonreductionist semantic theories can, of course, be used to handle this phenomenon (in a nonexplanatory way) by assigning different semantic values to an expression when it manifestly has different semantic properties. Thus a sentence with 'trailer' in it would be assigned different situations, or truth conditions, or extensions in possible worlds or markerese representations, depending on whether the dialect is American English or English English. Once again, this is accommodation, not explanation. The difference between CRS and the nonreductionist theories is that conceptual roles are, by their nature, system relative because they are functional entities and the semantic values of the nonreductionist theories are not.

It is worth emphasizing how important a matter this is. It is a banal feature of languages that the shape or sound of a word does not determine its meaning. Indeed, this point is sometimes described as "trivial semantic conventionalism," to distinguish it from more interesting claims. If no semantic theory could explain such a fact, semantics would be in trouble.

Perhaps it is worth mentioning the psychologism allegation again. Am I just demanding that semantics answer a question that belongs in the domain of, say, the psychology of language? Pretheoretically, the fact that one linguistic element can have different meanings in different languages would seem to be a clearly semantic phenomenon. I would think that the burden of argument would be on anyone who wanted to argue otherwise.

So CRS can explain the general fact that meaning is relative to representational system. Also, as pointed out in the last section, it promises to explain *particular* meaning differences. Since the difference in meaning of 'trailer' in English English (in which it means: movie preview) and American English is a matter of differences in the causal properties of the term, it is in

principle possible, according to CRS, to specify the factors that cause the difference in causal properties. By contrast, think of how a possible-worlds semanticist or a Katzian would go about explaining the difference. Nothing in such nonreductionist semantic theories would help.

The relativity of meaning to system of use is more fundamental to cognitive science than attention to examples such as 'trailer' indicates. Functional differences determine differences in the semantic (and syntactic) *categories* of representations—for example, the difference between the representational properties of *languagelike* and *picturelike* representations. This is especially important because there is reason to believe that many of our mental representations may actually be pictorial. None of the other semantic theories has a chance to explain the difference between the semantics of languagelike and picturelike representations.

Moreover, recall that syntactic category is as relative to system as semantic category. The relativity of syntactic category has the same explanation as the relativity of semantic category: syntax is functional too. If this isn't obvious, consider two processors that read English text: one reads odd-numbered characters, whereas the other reads even-numbered characters. One would read 'CDAOTG' as 'CAT', the other as 'DOG'. CRS allows a common explanation of an interesting fact—that both syntactic and semantic category are relative to system.

Further, CRS is important for avoiding misconceptions about concepts that are widespread in the psychological literature. The word 'concept' is used in psychology to denote a mental or physiological entity that expresses or represents a concept in the philosopher's sense of the term (in which concepts are abstract entities). The concept of a cat (in the psychologist's sense of the term) is a mental or physiological entity that expresses or represents cathood (much as the word 'cat' expresses or represents cathood). It is widely supposed in developmental psychology that mental images are probably children's concepts but that they could not be adult concepts. Piaget says:[67]

> The preconcepts of this level can be considered to be still half-way between the symbol and the concept proper. . . . [T]he preconcept involves the image and is partially determined by it, whereas the concept, precisely because of its generality, breaks away from the image. . . .

Another example: Premack (1982) argues that whereas the concepts of many lower animals are pictorial, the concepts of primates must be in part languagelike because pictorial concepts cannot express certain abstract ideas. For example, chimps can "match to sample" not only in cases where the sample is red and the correct multiple choice item is red, but also where the sample is AA and the choices are AB, BC, and BB. Here the correct choice is BB, and the common property is being a pair whose members are identical.

According to Premack, this requires a nonimagistic concept because the sample and target do not "resemble" one another. Another issue where this mistake (Which mistake? See the next paragraph) sometimes comes in is the issue of whether there is a "third code" more abstract than either languagelike or picturelike codes. The mistaken reasoning is that we have a nonlanguagelike code but that it could not be pictorial because pictorial representations could not have the kind of generality required of a concept.[68]

The doctrine that picturelike representations won't do for general or adult or primate concepts involves a conceptual error, one for which CRS is a corrective. CRS tells us that to be a concept of, say, dog, a mental representation must function in a certain way. Obviously, you can't tell how a certain representation functions by confining your attention to the representation alone, or to its "resemblances" to things in the world. You must know something about how the processors that act on it treat it. Thus a pictorial representation can express quite an abstract property, so long as the processors that act on it ignore the right specificities. To take a venerable example, a picture of an equilateral triangle can serve to represent triangles in general so long as the processors that act on it ignore the equality of the sides and angles. Similarly, a picture of a set of twins *could* represent or express the concept of a pair whose members are identical.

Note that I am not just pointing out that Piaget and Premack are the victims of "resemblance" theories of pictorial representation. The error I am pointing to is more fundamental in the sense that it includes the resemblance-theory error, plus a failure to see the shape of a positive doctrine—namely, that how or what a representation represents is a matter of more than the intrinsic properties of the representation or simple relational properties like "resemblance"; in particular, it is a matter of a complex relational property: how the representation functions.

What Is the Relation between Meaning and Mind/Brain?

How does the brain confer meaning on its representations?[69] Answer: By conferring the right causal roles on the representations. What is it for a person to grasp the meaning of a word? Answer: For a person to grasp the meaning of a word is for the word (or its standard mentalese associate) to have a certain causal role in his or her brain. How can it be that a person grasping an abstract object can propel the person (and his or her brain) to Hawaii? Answer: The difference between grasping a meaning and not grasping it is a difference in the causal role of entities in the person's brain, and differences in such causal roles can make for differences in behavior and the rewards that are contingent on behavior.

As before, the nonreductionist semantic theories can give superficial answers to the desideratum question. How does the brain grasp meanings? By grasping truth conditions or a denoted situation or a markerese structure.

But the question of how the brain grasps truth conditions or denoted situations or markerese structures is just as pressing as the original question.[70]

What Is the Relation between Autonomous and Inherited Meaning?

Recall the distinction (made in Desideratum 6) between autonomous and inherited meaning. Inherited meanings, like those of the linguistic expressions on this page, require translation or transliteration into the language of thought of a reader or hearer for their understanding. Autonomous meaning, the kind of meaning of the elements of the language of thought itself, requires no reading or hearing and thus no translation or transliteration in order to be understood. The questions I raised were: What is autonomous meaning? What is inherited meaning? What is the relation between autonomous and inherited meaning? For example, is one reducible to the other? Or are they both manifestations of a single type of meaning? Or are they unrelated phenomena with only a superficial resemblance?

The CRS answers to the first two questions are simple: autonomous meaning is conceptual role—and so is inherited meaning. (You will recall that using the notion of standard association, one can individuate conceptual roles of English as their standard associates in the internal system.) Further, the conceptual roles of external language are inherited from those of internal language. So inherited meaning is (surprise!) inherited from autonomous meaning.

The nonreductionist semantic theories, by contrast, have little to say about these matters. They *can* say that 'cat' and 'CAT' have the same semantic values; but as far as I can see, none of them have conceptual resources adequate to spell out any reasonable characterizations of autonomous and inherited meaning or say anything about whether one is reducible to the other.

Psychologism again: Is CRS supposed to be better for the purposes of psychology simply because it *contains* some psychological claims? Autonomous and inherited meaning are two categories of meaning (maybe even basic categories). It would be a surprise—*which itself would need explaining*—if no good theory of the nature of meaning could illuminate the issues I have been discussing about the relation between these two categories.

Indeed, once one sees the distinction between autonomous and inherited meaning, it is reasonable to ask of any theory of meaning *which* type of meaning it is intended to speak to. CRS speaks to both. Indeed, CRS explicates the difference between autonomous and inherited meaning without giving up a *unified* account of the two types of meaning. English inscriptions and utterances affect one another (via their effects on internal language) so as to give English expressions conceptual roles; and these conceptual roles are (at least on the simplified model I discussed) dependent on the conceptual roles of internal expressions.

Thus far, I have said little about causal theories of reference. Such theories, if they can be made to work, potentially have more to say about the relation between autonomous and inherited meaning than nonreductionist theories such as possible-worlds semantics, situation semantics, Davidsonian semantics, and Katzian semantics, because they can say something about the similarities and differences between the causal chains leading to 'cat' and 'CAT' that explains the differences and similarities between the two representations. But causal theories of reference cannot capture the aspect of meaning inside the head.[71] For example, they cannot capture the aspect of sameness in meanings of the sentences of me and my twin on Twin Earth (despite the difference in our causal chains outside our heads). From the point of view of a causal theory of reference, 'Hesperus = Hesperus' and 'Hesperus = Phosphorus' have the same semantic value.[72] Further, the theory that I am promoting can appropriate whatever successes causal theories of reference may have. Recall that CRS in the version I favor is part of a two-factor theory, the external factor of which can adopt aspects of a causal theory of reference account. In sum, causal theories of reference cannot accomplish the task I have set; and whatever they can accomplish can be appropriated by the two-factor version of CRS.

One final advantage of the CRS approach to the distinction between autonomous and inherited meaning is that it allows a theoretical approach to Searle's "Chinese Room" argument. With apologies to those who have heard this too many times: we are to imagine a monolingual English speaker who is placed in a room in a robot's head. He has a large library of instructions in English (the program) that tells him to push certain buttons (controlling outputs of the body) or write certain notes to himself (thus changing the "internal state" of the system) depending on what input lights are on and what notes he has written to himself earlier. The man never understands any Chinese, but nonetheless the robot he controls "speaks" excellent Chinese. Searle argues that since the man never understands Chinese, and since the robot paraphernalia adds no understanding, what we have is a Chinese simulator with no genuine Chinese understanding.

The most penetrating criticisms have focused on what Searle—anticipating the challenge—calls the systems reply. The systems reply says that since the system as a whole—man + library + room + robot body and control system—has the information processing characteristic of an intelligent Chinese speaker, we should take the whole system as understanding Chinese, even though the homunculus inside does not. The critics insist that the whole system does understand Chinese. (See Dennett 1983.) Searle has a clever reply. He tells the critics to just imagine the paraphernalia of the "system" *internalized,* as follows. First, instead of having the homunculus consult a library (the program), let him *memorize* the whole library. Second, let him memorize his notes instead of writing them down. Finally, instead

of having the homunculus inhabit a robot body, let him *use his own body*. That is, what we are to imagine in the new version is that the homunculus manipulates his own body in just the way he manipulated the robot body in the previous version. When he seems to be asking for the salt in Chinese, what he is really doing is thinking *in English* about what noises and gestures the program dictates that he should produce next.[73]

At this point, the issue seems to come down just to a matter of conflicting intuitions. The opponents say the man following the instructions does understand Chinese, Searle says he does not.[74] This is where CRS comes in. The trouble with the systems reply as so far discussed is that it contains no theoretical perspective on what it would be for the system's Chinese symbols to be meaningful for it in the way the symbols in the head of a normal Chinese speaker are meaningful for that person—it contains no perspective on autonomous meaning. CRS has an answer: what would give the symbols autonomous meaning is the right conceptual role. There is a complication that makes this point harder to see. Namely, there is a crucial ambiguity in Searle's statement of his examples. Is the robot system (and the later case in which the homunculus internalizes the program) supposed to be one in which the information processing of a normal Chinese speaker is *simulated*? Or is the information processing of a normal Chinese speaker actually *instantiated* or *emulated* in the system?[75] (I can simulate an Aristotelian physicist's information processing by figuring out what someone would think if, like Aristotle, he didn't distinguish average from instantaneous velocity; but I cannot instantiate or emulate this information processing—that is, have this type of information actually go on in me—because I cannot avoid seeing the distinction.) In the case of mere simulation, the information-processing point of view does not dictate that the system *does* understand anything. But in the emulation case—the one in which Chinese symbols are processed so that they have the same conceptual roles they have in a normal Chinese speaker.[76]—then CRS dictates that the robot does indeed understand Chinese. I think that what makes Searle's argument sound so convincing is that it is difficult to imagine a version of Searle's example that is a genuine instantiation or emulation rather than a mere simulation.[77] In sum, CRS allows one to see an important distinction that is not respected in the debate, and it gives those who are inclined toward functionalism a positive view about autonomous meaning so they can steer away from mere intuition-mongering.

What's the difference between Searle's argument and my argument in Block (1978)? To make a long story short, though our examples were similar, Searle's argument has a wider target, the symbol-manipulating view of the mind common in cognitive science. This view entails functionalism but is not entailed by it. My aim, by contrast, was mainly to argue that functional definitions constructed from commonsense psychology (by a Ramsification

procedure) carried a burden of proof. I argued that nothing of any substance had been said in their favor, and there was some reason to doubt them. Desiderata like the ones mentioned in this paper can be used to satisfy this burden of proof—for intentional states but not experiential states.

Compositionality

The points to be made about compositionality are very similar to points already made, so I will be brief.

According to CRS, it is sentences (and perhaps larger chunks of discourse) that embody hypotheses, claims, arguments, and the like, not subsentential elements. So, according to CRS, the semantic values of words and other subsentential elements are a matter of their contributions to the conceptual roles of sentences and supersentential elements. The conceptual role of 'and', for example, derives from such facts as that a commitment to rejecting '*p*' (in the absence of a commitment to accept '*p* and *q*') can lead (in certain circumstances) to a commitment to rejecting '*p* and *q*'. In this way, CRS explains why words have the conceptual roles they do by appeal to conceptual roles of sentences; thus the semantic values of words are seen to be a matter of their causal properties.

The nonreductionist theories do not and should not be regarded as aimed at this type of issue. They are concerned with what the relations among meanings of, say, words and sentences are, not with the issue of why those relations obtain.

What about indicator semantics and Gricean semantics? They, like CRS, take sentential and perhaps supersentential chunks as the basic semantic unit. And, like CRS, they can regard the meanings of words as their contributions to the semantic values of sentences. CRS has no advantage in this matter.

Narrow Meaning, Twin Earth, the Explanation of Behavior, and the Function from Context to Reference and Truth Conditions

The hard work of this section was done (or at any rate, attempted) in the desideratum on narrow meaning. I can be brief here, concentrating on objections and extensions.

What is narrow meaning? (Recall that CRS can do without the claim that narrow meaning is genuinely a kind of meaning, rather than a determinant of meaning.) Here, the comparison with the other theories looks quite different than with the other desiderata. CRS does have an answer—namely, conceptual role—and the other theories have no answer. But the other theories I've been mentioning are not *about* narrow meaning.

Why is narrow meaning relevant to the explanation of behavior, and why is it relevant in the same way for me and my twin? Taking the second

question first: since my twin and I are physically identical, all of our representations have exactly the same internal causal roles, and hence the same narrow meanings. But why is narrow meaning relevant to the explanation of behavior in the first place? To have an internal representation with a certain narrow meaning is to have a representation with certain likely inferential antecedents and consequents. Hence, to ascribe a narrow meaning is to ascribe a syndrome of causes and effects, including, in some cases, behavioral effects (or at least impulses in motor-output neurons). The reason my twin and I both jump is that we have representations with conceptual roles that have, as part of their syndrome of effects, jumping behavior. The reason that wide meaning is not as relevant to the explanation of behavior as is narrow meaning is that differences in wide meaning that do not involve differences in narrow meaning (e.g., the difference between me and my twin) do not cause behavioral differences.[78]

The CRS explanation of behavior may seem circular, hence trivial. How can I characterize a meaning functionally, in part in terms of a tendency for representations that have it to cause jumping, and then turn around and explain jumping by appeal to a representation's having this meaning? This is an objection of a well-known sort to explanation in terms of functionally individuated entities, and it has a familiar sort of rebuttal. 'Gene' is defined functionally in Mendelian genetics, in part in terms of effects on, for instance, hair color. 'Reinforcement' is defined in operant-conditioning circles in part in terms of effects on, for instance, bar-pressing. How, then, can one turn around and explain blonde hair in terms of genes, or bar-pressing in terms of history of reinforcement? Part of the answer is that one is not talking about a *single* effect, postulated ad hoc, but rather a complex web of interacting effects. A sickle-cell gene yields sickle-cell anemia in one circumstance (when paired with another sickle-cell gene) but resistance to malaria in another. When one postulates a gene on the basis of one effect, one can obtain converging evidence for it from other effects; and these effects enrich the functional characterization. If you give a rat Burpee Rat Chow (at 80% body weight[79]) contingent on bar-pressing, the rat's bar-pressing response normally increases in strength (on a variety of measures). So it is said that the Burpee Rat Chow is a reinforcer. Part of what makes this a nonempty claim is that one can get the rat to do all sorts of other things using Burpee Rat Chow or other reinforcers.

Second, and more importantly, a functionally individuated entity can, in principle, be identified by independent (usually physicalistic) means and the mechanism of its causal connection to the effects described. For example, a gene identified functionally via the methods of Mendelian genetics can be identified as a clump of DNA via the methods of molecular genetics. And the mechanism by which the gene produces phenotypic characteristics can be described biochemically. Similarly, the mechanism by which Burpee Rat

Chow affects behavior can (presumably) be characterized biologically, or perhaps even psychologically (in terms of the rat's information processing).

The application of the first point to CRS is obvious, but the application of the second is more problematic. The problem has to do with the type-token relation for mental representations. The hope is that there will be a stable physical realization (at least over short stretches of time) of, say, the representation 'CAT', which of course will be identifiable only by its functional role. Then, in principle, one could trace the causal links between this representation and behavior, just as the biochemist can in principle trace the mechanism by which a gene affects the phenotype.[80]

Let us now turn briefly to the matter of the essential indexical. 'I am in the path of danger', and 'Ned Block is in the path of danger' can have systematically different conceptual roles, depending on whether I know I am Ned Block (rather than, say, Napoleon). 'I', used by a speaker, differs systematically from the speaker's own name in its conceptual role, even though they refer to the same thing. Hence CRS assigns them different narrow meanings. Thus the thought I express with 'I' (or its internal associate) is different in narrow content from the thought I would have expressed were my name to have replaced 'I'. Thus, narrow meaning, as articulated by CRS, can be used to explicate a notion of thought *state* distinct from thought *object* that will serve the purpose for which Perry suggested this distinction.[81]

Similar points apply to the examples using names and natural kind terms mentioned in the desideratum on this subject. 'Cicero struts' and 'Tully struts' have different conceptual roles; so despite the fact that they have identical wide meanings, we can see why believing these different sentences could have different effects on other mental states and behavior.[82]

Let us now turn again to the determination of the function from context to reference and truth value. I argued in the section on meaning and reference/truth that conceptual role does determine this function. Take 'I', for example. If someone says "I am in danger," one can infer that the speaker has said, of himself, that he is in danger. In general, it is part of the conceptual role of 'I' that it refers to the producer of the token of 'I' (except in contexts such as quotation). However, there are other aspects of conceptual role that are relevant to, say, explanation of behavior, but not to determination of the function from context to reference and truth value. For example, one can infer from "I entirely fill such and such a spatiotemporal volume" to "You do not occupy this volume." But this inference does not seem relevant to the determination of the aforementioned function. Similar points apply to other types of terms. One can infer from 'water' to 'colorless' (or, at least to 'colorless if pure'); but this has little or nothing to do with determination of reference. I would still be referring to the same liquid even if I were under the impression that in its pure state it has a bluish tinge to it. Indeed, it may be that the aspect of conceptual role that determines the function from

context to reference is the same for all natural kind terms. My highly tentative conclusion is that the aspect of conceptual role that determines the function from context to reference and truth value is a small part of the conceptual roles relevant to the explanation of behavior and psychological state.

In conclusion: In this paper, I have not attempted to elaborate CRS, or supply any analyses of language from its perspective. Rather, I have tried to provide reason for suppressing the "put up or shut up" reflex that dogs talk of conceptual roles in the absence of identity conditions for them. My hope is that this theory will get more attention and that more detailed versions of it will allow us to evaluate its prospects better.[83]

Notes

1. Good sketches of the ideas of the representational theory of mind are to be found in Fodor (1981) and Lycan (1981). A more detailed treatment is provided in Pylyshyn (1984).

2. See Block (1983) for a discussion of this distinction and for references to the literature on this topic.

3. I hope my "inherited/autonomous" terminology won't make these questions seem trivial.

4. It is commitment to a priori truth (by which I mean truths for which there is no epistemic possibility of refutation) that really causes trouble for friends of analyticity—not our inability to come up with identity conditions for meaning. After all, no one has ever come up with satisfactory identity conditions for people or ships.

5. Perry (1977, 1979); Kaplan (unpublished).

6. A natural variant on the notion of narrow individuation that I described would require in addition that the same properties be attributed in the same way.

7. Note that the claim that narrow meaning is in the head, in this sense, is not incompatible with the idea that what it is for a word to have a certain narrow meaning is for it to express a concept, where concepts are taken to be abstract objects not locatable in space and time; in this respect, "in the head" is not an apt phrase.

8. Of course, one could define a referential notion of meaning that included narrow meaning and therefore better deserved to be called "wide." This would also result in a more intuitive treatment of vacuous reference. Since the main use I'll be making of the notion of wide meaning is to highlight narrow meaning, I'll stick with the simple definition I've introduced.

9. See Loar (1982), 279; White (1982); and Fodor (1985).

10. Cf. Field (1977).

11. This is a controversial reading of the lesson of Kripke's puzzle. I don't have the space here to describe either the puzzle or the conceptual role semantics solution.

12. White (1982) attempts to *define* a narrow meaning notion using such counterfactuals. But this seems misguided, since there is something shared by the twins *in virtue of which* the counterfactuals are true, and that seems a better candidate for narrow meaning.

13. See McGinn (1982), esp. 211-16, for arguments from the nature of representation to narrow content and meaning.

14. Ignore the problem that since we are made up largely of water, my twin and I can't be duplicates—fixes for this have been proposed by Putnam and Burge.

15. Burge (1979).

16. Actually, my position is that such a multidimensional gradient is needed for full-blooded narrow meaning, but not for the *part* of narrow meaning responsible for mapping contexts onto referents and truth conditions.

17. See Horwich (1982b). Here is the paradox of the preface: I write a book all of whose sentences I believe; nonetheless, I am sure that, being human, I have asserted at least one falsehood. Contradiction. Solution: I have a high degree of belief in each sentence in the book, but that is compatible with a high degree of belief in the falsity of their conjunction.

18. Variance is mean squared deviation from the mean.

19. Burge (1979). Burge constructs cases in which a man has a slight misunderstanding about how a word is used (e.g., he thinks you can have arthritis in the thigh). He then argues, persuasively, that a doppelgänger of this man in a language community in which 'arthritis' is standardly used to include rheumatoid inflammations of bones such as the thigh should not be regarded as meaning by 'arthritis' what we and our man mean by the word.

20. LePore and Loewer (1985) seem to object in this way to two-factor conceptual role semantics.

21. See Harman (1974, 1975, and 1982) and Sellars (1963, 1969, and 1974); see also Putnam (1979).

22. Field (1977, 1978).

23. See Churchland (1979), Loar (1981, 1982), Lycan (1981), McGinn (1982), and Schiffer (1981). Loar and Schiffer advocated conceptual role semantics only as a subsidiary semantic theory for the language of thought, if there happens to be one. The semantic theory they advocated for external language is a functionalized Gricean theory.

24. Woods (1977, 1978, and 1981).

25. Johnson-Laird (1977) and Miller and Johnson-Laird (1976).

26. Though in a paper given at the MIT Sloan Conference, 1984, Field suggests a view in which meaning and content are abandoned altogether. Field's 1977 and 1978 papers are quite skeptical about intersubjective comparisons of conceptual role—because of the collateral information problem. For that reason, he placed great weight on the referential component; recent skepticism about the referential component has led to skepticism about meaning and content altogether.

27. That is, the narrow aspect or determinant of meaning.

28. McGinn (1982) states the theory as assigning states of affairs to sentences. This leads LePore and Loewer (1985) to suppose that a two-factor theory must be more liberal than Davidsonian truth theory in allowing, in the external factor: 'Water is wet' is true $\leftrightarrow$ H_2O is wet. But a two-factor theorist *can* adopt Davidsonian truth theory for the external factor, even though demanding that the sentence on the right-hand side of the biconditional be a *translation* of the quoted sentence on the left-hand side is a stronger demand than necessary for the two-factor theorist.

29. For purposes of this discussion, I shall be ignoring pictorial internal representations.

30. Brain-writing, as everyone knows, is spelled in capital letters.

31. See Kahneman, Slovic, and Tversky (1982) and references therein for detailed studies of such fallacies.

32. Harman (1970) contrasts code-breaking views of language understanding with incorporation views. On the latter, understanding English is translation into a different language; whereas on the former, English is part of the language of thought (actually, a system of syntactic structures with English vocabulary items is part of the language of thought), so no translation is involved.

33. Harman (1982), 14.

34. See Loar (1982), 278-80, for a different slant on what is wrong with Harman's view. Loar takes the line that devices such as Harman's "normal context" and conceptual role in the minds of experts are ad hoc.

35. Johnson-Laird's reply (1978) to Fodor pretty much abandons this verificationist tendency in favor of a generalized conceptual role much like the idea I've been alluding to here.

36. Fodor (1978); reprinted in Fodor (1981), 211.

37. Dretske (1983), 88.

38. This is well argued by Stich (1983). (Although, as I think Sterelny (1985) shows, Stich deploys the wrong notion of "potential" in characterizing his functional roles.) Oddly, Stich considers mental representations, functionally individuated, without ever considering whether there is a distinction to be made between the aspect of functional role relevant to semantics and the aspect that might be called syntactic. (Indeed, these are in effect identified on p. 200.) This is a distinction we make with respect to English orthography. If someone writes the letter 'a' in an idiosyncratic way, we can identify it *functionally*, by the way it appears in words—e.g., it appears by itself, it appears in '*b*n*n*', in place of the asterisks, etc. At the same time, we can distinguish functionally between two uses of the same syntactic type, 'bank'.

39. Fodor (1985).

40. See Hills (1981), 18-19, for a dicussion of the two ways of talking about internal symbolism, and Harman (1973) for an application of the representational state version.

41. See Horwich (1982) for a discussion of this issue in another context.

42. Lycan (forthcoming) argues that God could tell us which worlds were the ones in which a sentence is true without telling us what the sentence means. I think he is right, but only for the reason mentioned in the text. God could indicate the possible worlds in a way that allows us to represent which ones they are without representing what they have in common in virtue of which they are the ones in which the sentence is true. See Lycan's paper for a discussion of indexicals and for references to the literature on this topic.

43. See also Loar (1982), 277.

44. This is Putnam's claim in an influential series of articles beginning with "The Analytic and the Synthetic" (1962); the few decompositional definitions he allows are those that, like 'bachelor = never-married adult male; involve a single "criterion." The idea is that the term 'bachelor' responds to only one "concern," and so there is no possibility that different concerns will "pull apart," creating a situation in which we will have to choose arbitrarily how the word is to apply. Putnam has also formulated a version of the argument given below against Fodor's innateness thesis.

45. Of course, it is not a particularly *new* story. Indeed, it is just what you would expect if you believed aspects of Quine and Kuhn, or if you accepted Lewis' "functional definition" story in "How to Define Theoretical Terms" (Lewis 1970). See Kuhn (1983) for semantic views quite close to those of conceptual role semantics.

46. I have heard it said that a conceptual role account of meaningfulness is much more plausible than a conceptual role account of particular meanings. This view is reminiscent of the cognitive theory of emotions that says that what makes a state an emotional state is a certain type of physiological arousal, but what makes such a state joy as opposed to anger is a difference in cognitive "overlay." The application of this idea to semantics cannot be evaluated in the absence of a suggestion as to what it is that accounts for the differences among meanings. Just one comment: in the case of experiential mental states, this type of view is less plausible than the reverse: that some sort of physiological state makes a state experiential, whereas functional *differences* are responsible for the difference between pain and the sensation of red.

47. These theories can often explain semantic defects in complex entities on the basis of the semantic properties of primitives. For example, Katzian semantics can explain why 'red idea' is semantically defective on the basis of the semantic values of 'red' and 'idea'. But Katzian semantics can give no answer to the question of what makes a primitive meaningful element meaningful. The Katzian accommodates the difference between 'red' and 'glub' by putting 'red' but not 'glub' in his dictionary. But it is not part of the theory to give an account of *why*.

48. This article is a jointly written pseudointerview in which the quoted material is put in Barwise's mouth (p. 51), but Perry continues the line of thought.

49. See Stalnaker (1984) for an attempt to solve this within the possible-worlds framework.

50. I derive this objection by analogy to a point made with regard to truth in Soames (1984), 426.

51. This may be Soames's view in the article mentioned in note 50, and I also see a tendency towards this view in Katz (1982), though Katz and Soames probably have different notions of necessity in mind.

52. Though, in the case of at least one version of the Gricean approach, not naturalistic.

53. See the statement of the theory in Schiffer (1982).

54. I used to think that the Fodor-Putnam multiple realizability arguments against physiological reductionism settled the matter. Their point, in essence, was that physiological reductionism was a chauvinist thesis in that, construed as a theory of the mind *simpliciter*, it would exclude intelligent machines or Martians. I now think that the best one is likely to get in the way of a theory of the mind will be a theory of the *human* mind. Such a theory will inevitably be chauvinist. The representational theory of mind that I am adopting here is a theory in that chauvinist tradition. What makes physiological reductionism look so bad is not simply that it is chauvinist—i.e., not just that there are merely *possible* creatures that share our intentional states without sharing our physiology—but rather that we do have promising theories of the human mind and that they are computational-representational (which is not to say that they are committed to the claim that the brain is a digital computer). If the scientific "essence" of intentional states is computational-representational, then it is not physiological—for the old multiple realizability reasons. So multiple realizability is the nub of the matter, but only because one chauvinist theory of the mind is multiply realizable in terms of another.

55. This comes through loud and clear in Searle (1980b).

56. Though in a draft of an article circulated in 1984, Schiffer rejects his earlier approach.

57. See Posner (1978) and Fodor (1982).

58. Sternberg (1969).

59. I can't possibly go into the details here. Dretske's view is couched in terms of the interesting notion of the *most specific information* that a tokening of a representation carries about a state of affairs.

60. Barwise (1984), 8.

61. The theory is sketched in Fodor (1983a, 1984) and expounded in detail in a widely circulated but as yet unpublished paper, "Psychosemantics" (see Loar [1983] for further comments on this paper), which Fodor is now saving for a book he is preparing of the same title. The reason I devote so much space to a largely unpublished account is that the problems with Fodor's account, together with Fodor's refutation of the Dretske-Stampe view, gives us an excellent picture of the type of problem faced by indicator semantics.

62. I am indebted to Paul Horwich here.

63. See, for example, the articles in the relevant section of Sober (1984).

64. There are disagreements about the *extent* of forces orthogonal to optimality. Lewontin and Gould, for example, are controversial in their insistence that the extent of such orthogonal forces is very great. (See their article in Sober [1984].) But this disagreement in the field should not obscure the important agreement mentioned in the text.

65. This issue can be discussed in terms quite distant from evolutionary biology. One example considered by Fodor is that when it comes to beliefs about poisons, false negatives are much more damaging than false positives. False positives ("This is a poison," said of something that is harmless) can cost you a meal, but false negatives can cost you your life. There are mechanisms in rats and even people that could perhaps be interpreted as inclining one to overattribute noxiousness to foods. Fodor insists that in such cases, one should *always* interpret the organisms as paying heed to low probabilities of very bad things rather than falsely ascribing high probabilities to the bad things. He sees this as a product of a principle of charity. The trouble with this reply is that this is not an a priori issue. If the mental sentence theory of belief is right, there is a difference between acting on a belief that p and acting on an estimate that, though p is unlikely, it would be terrible if true. Independent evidence could be marshalled in favor of one or another alternative. Further, even if Fodor's a priori assumption is right about our cognitive mechanisms, it is contingently right. If we come to understand how our cognitive

mechanisms work, perhaps we could build cognitive mechanisms that work otherwise. It would be a strange semantic theory that depends on such a highly contingent and perhaps quite alterable fact about the cognitive mechanisms that we happen to have. Such a semantic theory would not apply to robots who think, act, and talk almost exactly as we do, but, say, are built to overattribute poisonous qualities to foods on the basis of slim evidence. Will Fodor say we are barred by the logic of the concepts involved from building such a robot?

Another problem with Fodor's a prioristic method of handling these cases is that he is forced to adopt, ad hoc, *other* methods of handling other cases in which supposedly cognitive mechanisms don't aim at truth. In considering the possibility that our cognitive mechanisms are built to *repress* certain unpleasant truths, Fodor stipulates that such mechanisms are not cognitive. He is stuck with simply stipulating which mechanisms are cognitive and which are not.

66. There is a parallel problem in causal theories of reference that *seems* more tractable, but perhaps only because it is more familiar.

67. Quoted in Mandler (1983). On this issue as on many others, one finds glimmers of quite different views in Piaget. There are other passages where he seems to have some appreciation of the Berkeleyan point I make below. See also the discussion in Fodor (1975).

68. This is not the only argument for the third code. There are powerful empirical reasons for postulating a third code. See Potter, Valian, and Faulconer (1977) for both the good and the bad reasons for believing in a third code. Brison (n. d.) (1984) contains an excellent rebuttal of arguments for a third code that make this (and other) mistakes. See also Kolers and Brison (1984).

69. Recall that I am ignoring the mind, concentrating on the brain.

70. The issue of psychologism naturally comes up with respect to this issue, but I have already answered it a number of times.

71. Unless they include in their causal chains the causal roles inside the head, in which case they include CRS itself.

72. Field (1977), 390.

73. This example is similar to ones described in Block (1978, 1981).

74. See the replies in the issue of *Behavioral and Brain Sciences* in which Searle's article appeared and the interchange between Searle and Dennett in *New York Review of Books* (Searle 1983b; Dennett 1983).

75. See Block (1980, 1981).

76. At the appropriate level of abstraction, of course. In this case, as in others I have mentioned, identity of conceptual role is compatible with a variety of causal differences.

77. The only reply I've seen that contains a glimmer of the CRS reply is Haugeland's in the BBS issue just mentioned (Haugeland 1980).

78. Burge (1984) objects that this use of 'behavior' begs the question in favor of individualistic accounts, behavioral ascriptions often being nonindividualistic. I agree that ordinary behavior descriptions are nonindividualistic; I would argue along the lines suggested in Desideratum 8 that an important line of work in cognitive psychology *is* individualistic.

79. To make sure it is hungry—an explanation avoided by most of those who condition rats.

80. Actually, I think there is less of a problem here than meets the eye. Letters of the alphabet are individuated functionally—that is why we recognize shapes that we have never seen before as *A*'s. But what allows us to do this is some degree of stability in the shapes of other letters. It is hard to see how there could fail to be some analogous story about how the brain works—if representationalism is true.

81. This point is similar to the one made by Lycan (1981), (See also Dennett [1982].) However, Lycan somehow sees this point as an argument for the internal sentence story (the conceptual role semantics comes in almost incidentally). I talk about thoughts rather than beliefs because the representationalist story is more plausible for occurrent mental states. As

many commentators have pointed out, one can ascribe a belief if it follows in a simple way from what a person has explicitly thought, even if the belief ascribed has never actually occurred to the person. See Fodor (n.d.).

82. On Kripke's puzzle: since 'Londres' and 'London' have different conceptual roles, it is a mistake to accept Kripke's translation principle. In particular, from the fact that Pierre croit que Londres est jolie, we should not conclude that Pierre believes London is pretty—if the content of his belief is given by 'London is pretty'. Lycan (1981) and McGinn (1982) have interesting discussions of the conceptual role semantics response to Kripke's puzzle, but neither pinpoint the translation principle as the culprit.

83. I am grateful to the John Simon Guggenheim Memorial Foundation and the Center for the Study of Language and Information for support while writing this paper. I would like to thank Michael Bratman, Martin Davies, Hartry Field, Jerry Fodor, Gilbert Harman, Paul Horwich, David Israel, Phil Johnson-Laird, Jerry Katz, Brian Loar, Bill Lycan, and Georges Rey for their helpful comments on earlier drafts.

References

Barwise, John. 1984. *The Situation in Logic—I.* Technical report CSLI-84-2, Stanford University, March. Paper presented at International Congress on Logic and Philosophy of Science, Salzburg, July 1983.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes.* Cambridge, Mass.

Barwise, Jon, and John Perry. 1984. *Shifting Situations and Shaken Attitudes.* Research report CSLI-84-13, Stanford University, August. To appear as a reply to critics in a special issue of *Linguistics and Philosophy* devoted to situation semantics.

Block, Ned. 1978. "Troubles with Functionalism." In *Perception and Cognition: Issues in the Foundations of Psychology,* edited by C. W. Savage. Minneapolis.

Block, Ned. 1980. "What Intuitions about Homunculi Do Not Show." *Behavioral and Brain Sciences* 3: 425-26.

Block, Ned. 1981. "Psychologism and Behaviorism." *Philosophical Review* 90: 5-43.

Block, Ned. 1983. "Mental Pictures and Cognitive Science." *Philosophical Review* 92: 499-541. Reprinted in *The Philosophers' Annual* 6 (1984), edited by P. Grim et al.

Block, Ned, and Sylvain Bromberger. 1980. "States' Rights." *The Behavioral and Brain Sciences* 3.

Boyd, Richard. 1979. "Metaphor and Theory Change." In *Metaphor and Thought,* edited by Andrew Ortony. Cambridge.

Brison, Susan J. n.d. "Do We Think in Mentalese?" Forthcoming.

Burge, Tyler. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4:73-121.

Burge, Tyler. 1984. "Individualism and Psychology." Paper presented at Cognitive Science Conference, Massachusetts Institute of Technology. Stephen Stich and Ned Block were the respondents. This paper will appear in the Philosophical Review.

Churchland, Paul M. 1979. *Scientific Realism and the Plasticity of Mind.* Cambridge.

Churchland, P. M. and P. S. Churchland. 1983. "Content—Semantic and Information-Theoretic." *Behavioral and Brain Sciences* 6: 67-78.

Davidson, Donald. 1984. *Truth and Interpretation.* Oxford.

Dennett, Daniel C. 1982. "Beyond Belief." In *Thought and Object: Essays on Intentionality,* edited by Andrew Woodfield. Oxford.

Dennett, Dan. 1983. "The Myth of the Computer: An Exchange." *New York Review of Books* (June 14). This contains a letter from Dennett criticizing Searle (1983b).

Devitt, Michael. 1981. *Designation.* New York.

Dretske, Fred. 1981. *Knowledge and the Flow of Information.* Cambridge, Mass.

Dretske, Fred I. 1983. "Why Information?" *Behavioral and Brain Sciences* 6: 82-89. This is Dretske's reply to critics.

Field, Hartry. 1972. "Tarski's Theory of Truth." *Journal of Philosophy* 69: 347-75.
Field, Hartry. 1977. "Logic, Meaning, and Conceptual Role." *Journal of Philosophy* 74: 379-409.
Field, Hartry. 1978. "Mental Representation." *Erkentniss* 13: 9-61.
Field, Hartry. 1980. *Science without Numbers: A Defense of Nominalism.* Oxford.
Fodor, J. A. 1974. "Special Sciences." *Synthese* 28: 77-115. Reprinted as part of Fodor (1975) and in my *Readings in Philosophy of Psychology,* vol. 1.
Fodor, J. A. 1975. *The Language of Thought.* New York.
Fodor, J. A. 1978. "Tom Swift and His Procedural Grandmother." *Cognition* 6: 229-247.
Fodor, J. A. 1981. *RePresentations.* Cambridge, Mass.
Fodor, J. A. 1982. "Cognitive Science and the Twin-Earth Problem." *Notre Dame Journal of Formal Logic* 23: 98-118.
Fodor, J. A. 1983a. "A Reply to Brian Loar's 'Must Beliefs Be Sentences?' " In *PSA 1982,* vol. 2, edited by Peter Asquith and Thomas Nickles. Ann Arbor.
Fodor, J. A. 1983b. *The Modularity of Mind.* Cambridge, Mass.
Fodor, J. A. 1984. "Semantics, Wisconsin Style." *Synthese* 59: 1-20. This is primarily an attack on Dretske and Stampe, but it does contain a brief exposition of Fodor's own theory.
Fodor, J. A. 1985. "Banish DisContent." In *Proceedings of the 1984 Thyssen Conference,* edited by Jeremy Butterfield. Cambridge.
Fodor, J. A. n.d. *Psychosemantics.* Forthcoming.
Harman, Gilbert. 1970. "Language Learning." *Nous* 4: 33-43. Reprinted in *Readings in Philosophy of Psychology,* vol. 2, edited by Ned Block. Cambridge, Mass., 1981.
Harman, Gilbert. 1973. *Thought.* Princton, N.J.
Harman, Gilbert. 1974. "Meaning and Semantics." In *Semantics and Philosophy,* edited by M. K. Munitz and Peter Unger. New York.
Harman, Gilbert. 1975. "Language, Thought and Communication." In *Language, Mind and Knowledge,* edited by K. Gunderson.
Harman, Gilbert. 1982. "Conceptual Role Semantics." *Notre Dame Journal of Formal Logice,* 23: 242-56.
Haugeland, John. 1980. "Programs, Causal Powers, and Intentionality." *Behavioral and Brain Sciences* 3: 432-33.
Hills, David. 1981. "Mental Representations and Languages of Thought." In *Readings in Philosophy of Psychology,* vol. 2, edited by Ned Block. Cambridge, Mass.
Horwich, Paul. 1982a. "Three Forms of Realism." *Synthese* 51: 181-201.
Horwich, Paul. 1982b. *Probability and Evidence.* Cambridge.
Johnson-Laird, P. N. 1977. "Procedural Semantics." *Cognition* 5: 189-214.
Johnson-Laird, P. N. 1978. "What's Wrong with Grandma's Guide to Procedural Semantics: A Reply to Jerry Fodor." *Cognition* 6: 241-61.
Kahneman, D. P. Slovic, and A. Tversky. 1982. *Judgement under Uncertainty: Heuristics and Biases.* Cambridge.
Kaplan, David. n.d. "Demonstratives." Circulated in mimeograph form since 1977 and the subject of the 1980 John Locke Lectures.
Katz, Jerrold J. 1972. *Semantic Theory.* New York.
Katz, J. J. 1982. *Language and Other Abstract Objects.* Totowa, N.J.
Kolers, Paul A., and Susan J. Brison. 1984. "On Pictures, Words and Their Mental Representations." *Journal of Verbal Learning and Verbal Behavior* 23: 105-13.
Kripke, Saul. 1979. "A Puzzle about Belief." In *Meaning and Use,* edited by A. Margalit. Dordrecht.
Kuhn, Thomas S. 1983. "Commensurability, Comparability, Communicability," In *PSA 1982,* vol. 2, edited by Peter Asquith and Thomas Nickles. Ann Arbor.
LePore, E., and B. Loewer. 1985. "Dual Aspect Semantics." In a festschrift for Donald Davidson, edited by E. LePore. I sent the authors a number of criticisms of a draft of this paper; I have no idea whether the version to be published retains the points I criticize here.

Lewis, David. 1970. "How to Define Theoretical Terms." *Journal of Philosophy* 67: 427-46.
Loar, Brian. 1981. *Mind and Meaning.* Cambridge.
Loar, Brian. 1982. "Conceptual Role and Truth Conditions." *Notre Dame Journal of Formal Logic* 23: 272-83.
Loar, Brian. 1983. "Must Beliefs Be Sentences?" In *PSA 1982,* vol. 2, edited by Peter Asquith and Thomas Nickles. Ann Arbor.
Lycan, W. 1981. "Toward a Homuncular Theory of Believing." *Cognition and Brain Theory* 4:139-59.
Lycan, William G. "Semantic Competence and Truth Conditions." Unpublished manuscript.
Mandler, Jean M. 1983. "Representation." In *Cognitive Development,* edited by P. Mussen. Vol. 3 of *Manual of Child Psychology.* New York.
Marr, David. 1982. *Vision.* San Fransciso.
McGinn, Colin. 1982. "The Structure of Content." In *Thought and Object,* edited by Andrew Woodfield. Oxford.
Miller, G. A., and P. N. Johnson-Laird. 1976. *Language and Perception.* Cambridge, Mass.
Perry, John. 1977. "Frege on Demonstratives." *Philosophical Review* 86: 474-97.
Perry, J. 1979. "The Problem of the Essential Indexical." *Nous* 13: 3-21.
Posner, M. I. 1978. *Chronometric Explorations of Mind.* Hillsdale, N. J.
Potter, M. C., V. V. Valian, and B. A. Faulconer. 1977. "Representation of a Sentence and Its Pragmatic Implications: Verbal, Imagistic, or Abstract?" *Journal of Verbal Learning and Verbal Behavior* 16: 1-12.
Premack, David. 1983. "The Codes of Man and Beast." *Behavioral and Brain Sciences* 6: 125-37.
Putnam, Hilary. 1962. "The Analytic and the Synthetic." In *Mind, Language and Reality,* edited by Hilary Putnam. Cambridge.
Putnam, Hilary. 1975. "The Meaning of 'Meaning'." In *Language, Mind, and Knowledge,* edited by K. Gunderson. Minneapolis. Also in Putnam's *Mind, Language and Reality.*
Putnam, Hilary. 1979. "Reference and Understanding." In *Meaning and Use,* edited by Avishai Margalit. Dordrecht.
Putnam, Hilary. 1983. "Computational Psychology and Interpretation Theory." In *Realism and Reason,* edited by Hilary Putnam. Cambridge.
Pylyshyn, Zenon. 1984. *Computation and Cognition.* Cambridge, Mass.
Schiffer, Stephen. 1981. "Truth and the Theory of Content." In *Meaning and Understanding,* edited by H. Parret. Berlin.
Schiffer, Stephen. 1982. "Intention-Based Semantics." *Notre Dame Journal of Formal Logic,* 23: 119-59.
Searle, John. 1980a. "Minds, Brains and Programs." *Behavioral and Brain Sciences* 3: 417-24.
Searle, John. 1980b. Searle's reply to critics of "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3: 450-57.
Searle, John. 1983a. *Intentionality: An Essay in the Philosophy of Mind.* Cambridge.
Searle, John. 1983b. "The Myth of the Computer." *New York Review of Books* (June 14).
Searle, John. 1984. "Intentionality and Its Place in Nature." *Synthese* 61:3-16.
Sellars, Wilfrid. 1963. *Science, Perception and Reality.* London. See "Empiricism and the Philosophy of Mind" and "Some Reflections on Language Games."
Sellars, Wilfrid. 1969. "Language as Thought and as Communication." *Philosophy and Phenomenological Research* 29: 506-27.
Sellars, Wilfrid. 1974. "Meaning as Functional Classification." *Synthese* 27: 417-27.
Shoemaker, Sydney, 1984. *Identity, Cause, and Mind.* Cambridge.
Soames, Scott. 1984. "What Is a Theory of Truth?" *Journal of Philosophy* 81: 411-29.
Sober, Elliot. 1984. *Conceptual Issues in Evolutionary Biology.* Cambridge, Mass.
Stalnaker, Robert C. 1984. *Inquiry.* Cambridge, Mass.

Stampe, Dennis W. 1977. "Toward a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy* 2: 42-63.

Sterelny, Kim. "Is Semantics Necessary? Stephen Stich's Case Against Belief." Forthcoming. To appear in *The Australasian Journal of Philosophy.*

Sternberg, S. 1969. "Memory Scanning: Mental Processes Revealed by Reaction Time Experiments." *American Scientist* 57: 421-57.

Stich, Stephen. 1983. *The Case Against Belief.* Cambridge, Mass.

White, Stephen L. 1982. "Partial Character and the Language of Thought." *Pacific Philosophical Quarterly* 63: 347-65.

Woods, William. 1977. "Meaning and Machines." In *Proceedings of the International Conference on Computational Linguistics,* edited by A. Zampoli, Florence.

Woods, William. 1978. *Semantics and Quantification in Natural Language Question Answering.* Technical report 3687. Cambridge, Mass.

Woods, William. 1981. "Procedural Semantics as a Theory of Meaning." In *Elements of Discourse Understanding,* edited by A. Joshi, B. Webber, and I. Sag. Cambridge.