INTENTIONALITY AND COGNITIVISM

by

Ronald Albert McClamrock

B.A. University of Washington
(1980)

Submitted to the Department of
Linguistics and Philosophy
in Partial Fulfillment of the
Requirements of the Degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1984

Copyright Ronald A. McClamrock, 1984

The author hereby grants to M.I.T. permission to reproduce and to distribute copies of this thesis document in whole or in part.

Signature of Author: _____
Department of Linguistics and
Philosophy, May 4, 1984

Certified by: _____
Ned Block
Thesis Supervisor

Certified by: _____
Jerry Fodor
Thesis Supervisor

Accepted by: _____
Richard Cartwright
Chairman, Departmental Committee on
Graduate Students

ARCHIVES
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 27 1984

LIBRARIES

ABSTRACT

Cognitivism, according to which the mind is to be characterized in terms of the brain's information processing structure, has recently gained some prominence in philosophy of mind. A contemporary version of the representationalism of Locke and Descartes, this outlook is tied to and motivated by the widespread use of computational models in both cognitive psychology and artificial intelligence. My thesis explores some problems for this cognitivistic outlook which arise from a consideration of the intentional or semantic properties of mental states. The thesis consists of three independent parts:

In the first section, I assess John Searle's claim that the intentionality of brain states depends essentially on their biochemical rather than computational properties. I argue that his account depends on confusing cognitivism with behaviorism, "qualitative content" with intentional properties, and chemical properties with the constraints they place on interaction with the world. Furthermore, his treatment of the semantics of mental states either fails to answer the pertinent questions or else answers them incorrectly.

In the second, I discuss two related problems for the view that meaning is determined by cognitive structure. First, in the context of the familiar "twin-earth" examples, I argue against Tyler Burge's claim that natural kind terms require fundamentally different treatment than explicit indexicals like "I" and "now". Second, I evaluate Hilary Putnam's suggestion that any attempt to factor extension out of meaning will leave one with no reasonable criterion for sameness of meaning. I offer some criticism of Jerry Fodor's "denotational semantics" response to Putnam's problem, and suggest an alternative approach.

The third section addresses the relationship of the cognitivist view of intentionality to that offered by Husserlian phenomenology. Hubert Dreyfus, among others, has pointed out important parallels between the "methodological solipsism" of cognitivism and Husserl's "bracketing", and has used these parallels in arguing that putative problems for the Husserlian account also impugn the cognitivist's position. I contend that he exaggerates the problems for a Husserlian account, and that such difficulties as he does uncover may be avoided by cognitivism.

Advisors: Professors Ned Block and Jerry Fodor.

ACKNOWLEDGEMENTS

I would first of all like to thank my thesis advisors, Ned Block and Jerry Fodor. Their careful criticism of my work, quiding comments, and general intellectual stimulation has been invaluable. I have also benefitted greatly from the opportunity to discuss my work -- and philosophy in general -- with my fellow students. I would particularly like to thank Ken Albert, Jay Lebed, Greg Smith, and Thomas Uebel for the many profitable and enjoyable discussions I've been fortunate enough to share with them.

This work was supported in part by a fellowship from the M.I.T. Center for Cognitive Science, from a grant from the Alfred P. Sloan Foundation.

Table of contents:

PART 1:

THE CHEMISTRY

OF

INTRINSIC INTENTIONALITY

> It must be confessed, moreover, that perception and that which depends on it are inexplicable by mechanical causes, that is, by figures and motions. And, supposing that there were a machine so constructed as to think, feel and have perception, we could conceive of it as enlarged and yet preserving the same proportions, so that we might enter it as into a mill. And this granted, we should only find on visiting it, pieces which push one against another, be never anything by which to explain a perception. This must be sought for, therefore, in the simple substance and not in the composite or machine. [1]
>
> - Leibniz, The Monadology

I. Introduction

Intentionality is in again; it has resurfaced as a leading contender for that special something which we have but that the best of machines might yet lack. This "directedness upon an object" and "having within itself a content" which Brentano proposed as the identifying mark of the mental has, after a bit of a layoff, returned to the job of isolating our own mentality from pretenders to the title. What is this strange feature of "of-ness" had by protoplasm but not silicon, where does it come from, and why is it important?

In what follows, I'll examine a few of the issues

surrounding this topic by inspecting some of the positions and arguments of John Searle, one of the main combatants in the current discussions; emphasizing in particular the presentation given in his paper "Minds, Brains, and Programs." [2] Searle takes his central point in this article to be the refutation of what he calls "strong AI": the view that having a mind (intelligence, beliefs, etc.) is just a matter of embodying a certain sort of computer program -- of engaging in a particular kind of activity which can be "defined in terms of computational operations on purely formally defined elements." [3] The basic method of argument consists in taking the reader through a series of gedankenexperiments, each of which purports to present a candidate which some version of the "strong AI" view would count as among the mindful. We are then implored to accede in the intuition that the cases under consideration provide counter-examples to the view in question; for in each (we are assured), something is missing: intentionality -- the juice of meaning. I'll be avoiding much discussion of Searle's specific examples; rather than simply constructing imaginary cases and then reading off our unexamined intuitions from them -- a blatant sort of "intuition pump" strategy -- I'll try to bring out what seem to be the underlying principles which might be leading Searle to make the sorts of intuitive judgements about the examples that he does, and consider their plausibility and potential justification -- occasionally

through the use of additional examples.

It's important to get clear from the start on the notion of instantiating the same computer program which plays a central role in Searle's arguments, as it's instantiating the same program (of whatever kind you like) as, say, I do, which Searle is intent on rejecting as being itself sufficient for mentality. Now it sometimes looks like Searle is using this notion in a fairly weak sense; that is, in the sense of something roughly like that of computing the same function. Such a reading would certainly mesh well with Searle's preoccupation with the Turing test (especially in the "Chinese room" example, where the emphasis is centrally on the preservation of input/output relationships), and would support his claim to sameness of program in the cases -- like the Chinese room -- which get the strongest intuitive "no" vote on the presence of mentality. Indeed, it sometimes looks as though all Searle is actually concerned with is refuting something like the Turing test as providing a criterion for the mental; but if this is it, there would seem to be better arguments around for the same conclusion. [4] As far as this extremely weak notion of sameness of program goes -- a notion which is essentially a behavioristic one -- I for one am perfectly willing to accept his cases as counterexamples to "strong AI"; two things having the same input/output relations -- of whatever sort you like -- is perfectly compatible with one having a mind and the other

not.

But I think that it's clear there's a straightforward and natural notion of sameness of program which is much stronger than this, and which is the one Searle is actually concerned with. The weaker notion is the one relevant to a behavioristic account of mind; but it's this stronger notion which is central to a position which Searle is clearly interested in refuting as well. This is the outlook of cognitivism: (very roughly) the view that there is a true cognitive theory of the way we process information, and that it is in virtue of our falling under such a theoretical description that we have mentality. Now a rough-and-ready characterization of this stronger notion of sameness of program isn't too hard to get: it requires not only computing the same function, but doing it in the same way. Of course this is admittedly quite vague, and "in the same way" of course needs to be fleshed out a bit; surely the cognitivist doesn't want it to come to "by using the same physical mechinisms." Nonetheless, the intuitive idea for the cognitivist is, I think, pretty clear: to process information in the same way is to instantiate a sort of flow chart describing the information flow between primitive "black box" processors which manipulate representations and pass them among themselves. However this is to be filled out, this much seems clear: It is the stronger notion which Searle is actually concerned in refuting as providing a criterion for the mental; and given this, instantiating a

program which I instantiate and in virtue of which I have mentality will involve the preservation of some plausibly significant aspect of my internal structure.

Now Searle claims throughout his discussion of this issue that preservation of our brain's chemical structure (or what would presumably entail that -- its microphysical structure) is sufficient for the preservation of mentality. Of course for Searle, sameness of chemical structure isn't in any sense a functional property; rather, it requires sameness of underlying stuff. As far as preservation of mentality via sameness of program rather than stuff goes, Searle considers two main candidates which we might call cognitive and neural equivalence. The former is just what you'd think: for something the be cognitively equivalent to me, it must accomplish its information processing in the same way I do. That is, pick your favorite (true) theory of the way which I process information of the general sort that cognitive psychologists would like to provide, in terms of computations defined over representations accomplished by the primitive "black boxes" and their interrelations; then cognitive equivalence requires that this theory be true of the entity in question as well.

As for what I am calling neural equivalence, the matter is slightly less straightforward. In the sense intended here, neural equivalence to me does not require that the entity in question have a neural system which falls

under the same sorts of physico-chemical descriptions as does mine; rather, it requires that the entity instantiate some description of my neural system which characterizes neurons by their functional relations with each other. Such a description might characterize the functional relationships between neural firings, kinds of synaptic transmissions, etc., but would not determine the sorts of physical and chemical processes which would have to underlie such activity. Put slightly differently, we might think of neural equivalence in the intended sense as involving falling under the same sort of "black box" theory as is involved in the notion of cognitive equivalence; the difference being that in this case the primitive black boxes would correspond to individual neurons. In what follows, however, I'll confine my attention to the question of cognitive equivalence, and will use the more general "computational equivalence" to refer only to this case. Most (but not all) of what I say applies to this functional neural isomorphism as well; but surely it's the issue of cognitive equivalence which is more interesting, if for no other reason than it would seem to be central to the ideology of cognitive science in a way that neural equivalence is not central to any scientific ideology. [5]

We might then put what would seem to be a central pillar in a view such as Searle's in the following way: Neither cognitive nor neural equivalence with a normal human being is itself sufficient for mentality; something could

have either or both of these and yet fail to have the mark of the mental -- intrinsic intentionality. It is, on Searle's account, the sort of preservation of structure which is involved in cognitive and neural equivalence which is to be picked out by the notion of instantiating the same computer program; and it is just this sort of preservation of structure which is held by Searle to be itself insufficient for mentality.

Before moving on to the substantive claims and arguments offered, let me note here that the sort of alleged problem which Searle is trying to put forth for the cognitivist might also be stated, instead of in terms of "strong AI" and computer programs, in terms of the doctrine of psychofunctionalism and the problems of liberalism and chauvinism for functionalist accounts of the mental in general. [6] Psychofunctionalism (roughly, the view that to have a mind is -- put in the terminology at hand -- just to be cognitively equivalent, more or less, to a normal human being) is, viewed one way, a strategy for revising a more general functionalist kind of view in order to escape problems of liberalism; i.e., counting among the mindful candidates which should clearly be ruled out. Now Psychofunctionalism considered as specifying the nature of the mental is surely chauvinistic in ruling out entities which may not share our particular kind of cognitive structure, but may have a good claim to mentality nonetheless. But the cognitivist (or "strong AI" partisan),

by only purporting to give a sufficient condition for mentality, avoids this problem with chauvinism, which at the same time allowing himself the option of suggesting as a sufficient condition (at least for starters) a specification of internal structure which for the Psychofunctionalist (who wants necessary and sufficient conditions) would be absurdly chauvinistic; e.g., perfect computational equivalence to some particular human being. Given this, we might restate Searle's objection to cognitivism as this: Even such an absurdly chauvinistic criterion as this is still too liberal as well.

II. Intentionality, Consciousness, and Brains in Vats

Behind the intuition pump and the Chinese room, it is the issue of content on which Searle's position rests. For, he claims, at least some mental states are essentially contentful; or as he would put it, intrinsically intentional. But surely, he adds, formal symbol mainipulations are not by themselves meaningful at all. Thus, the reason that program is itself insufficient for mentality:

> Because the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even symbol manipulations, since the symbols don't symbolize anything. In the linguistic jargon, they have a syntax but no semantics. [7]

One common response to this challenge of Searle's to say where apparently uninterpreted symbols might get some

sort of meaning is to hold that they get their interpretation from their causal interactions with the ouside world. As William Lycan puts it in his commentary on Searle's acticle:

> ...no computer has or could have intentional states merely in virtue of performing sytactic operations on formally characterized elements.... Our brain states do not have the contents they do just in virtue of having their purely formal properties either; a brain state described "sytactically" has no meaning or content of its own. In virtue of what, then, do brain states (or mental states however construed) have the meanings that they do? Recent theory advises that the content of a mental representation is not determined within the owner's head; rather, it is determined in part by the objects in the environment that actually figure in the representation's etiology and in part by social and contextual factors of several other sorts. [7]

Thus (on this line) a substantial part of Searle's point is granted: formal symbol manipulation of whatever sort you like is not itself sufficient for intentionality; you also need these formal structures to have the right kind of causal (perhaps contextual in general) relationships to the outside world to get them interpreted. One might say that the interpretation and hence the intentionality of the formal system's states and representations comes from its dasein. [9] But Searle will have none of this. For, he claims, "that the internal operations of the brain are causally sufficient for mental phenomena is fairly evident from what we do know." [10] What we know that makes this evident, he goes on to say, are such things as that (with respect to some visual experience of a tree) "I could be

having exactly that visual experience even if there were no tree there, provided only that something was going on in my brain sufficient to produce the experience." [11] Perhaps unsurprisingly, we find lurking behind this "knowledge" everyone's favorite twentieth-century version of Cartesian doubt: "If I were a brain in a vat I could have exactly the same mental states I have now; it is just the most of them would be false or otherwise unsatisfied." [12]

Note, however, that Searle needs here (at least) what might be called the "strong" brain-in-a-vat intuition: not only might I be a brain in a vat now and have these very same mental (and intentional) states, but I might still have had these very same intentional states had I never been anything but a brain in a vat. Since for Searle, the fact that our mental states have the content they do is, as it were, a purely internal (to the brain) matter (as he says, a matter of the brain's biochemistry), these states must then have their content even if they are totally divorced from any sort of causal connection with the world through which they might manage to squeeze a little content. We might then imagine that rather than having the infamous mad scientist kidnap someone and remove her brain for his heinous experiment in semantics, we instead have the whole mess -- brain, nutrient bath, and "evil demon" computer -- materialize out in space from the random motions of particles. [13] We can thus (at least try to) avoid the possibility that the brain is somehow

"coasting on leftover dasein"; the possibility of any obvious sort of teleological or evolutionary story about content would thereby appear to be ruled out as well. [14]

If we were to grant Searle all this, we would then seem to be faced with the following problem: Brains don't need dasein for their intentionality; some kind of content is guaranteed by their internal operation. But surely, we are urged, no such thing is true of a formal symbol manipulator as such; the only sort of interpretation which its states and symbols get must come from the outside. As Georges Rey has pointed out, a computer might well run through exactly the same computational states on two days, but have its inputs and outputs be interpreted on one day as being about, say, the SALT talks, and on the next day as being about a chess game; "It's just that on Wednesday the punches in the cards are interpreted (say, by Carter) to refer to Brezhnev, Vienna, and 100-megaton bombs; and on Thursday the very same punches are interpreted (say, by Spassky) to refer to moves and pieces in chess." [15] As far as the computer is concerned, there just isn't any difference. Thus (as Searle would have it), whereas brains have intrinsic intentionality, formal symbol manipulators as such are only the objects of observer-relative ascriptions of intentionality: "a manner of speaking about the intentionality of the observers", which is "always dependent on the intrinsic intentionality of the observers." [16] And

as far as mentality is concerned, the latter sort of intentionality is (on Searle's line) no intentionality at all:

> ...the mental-nonmental distinction cannot be just in the eye of the beholder but it must be intrinsic to the system.... [17] There are not two kinds of intentional mental states; there is only one kind, those that have intrinsic intentionality; but there are ascriptions of intentionality [i.e. the observer-relative ones] in which the ascription does not ascribe intrinsic intentionality to the subject of the ascription. [18]

t this point, the following question needs to be asked: What's behind the intuition that the states of the brain in a vat have some kind of intrinsic inner content? Surely one central underlying intuition for Searle here is that this sort of content either is or is fundamentally derivative from the content of consciousness. Now if this is where the intrinsic intentionality of the brain's representational states is to come from, then at least Searle's in good company. The intentionality of consciousness has had an illustrious history of demarcating the realm of the mental; and the phenomenological tradition as a whole, having taken the issue of intentionality for its own, would apparently point us toward just this sort of view, on which the connection between intentionality and consciouness is -- to say the least -- intimate. As Husserl writes:

> What forms the materials into intentional experiences and brings in the specific element of

> intentionality is the same as that which gives its specific meaning to our use of the term "consciousness", in accordance with which consciousness points eo ipso to some thing of which it is the consciousness.... Consciousness is just consciousness "of" something; it is its essential nature to conceal "meaning" within itself.... [19]

Or as Sartre more simply puts it: "Indeed, consciousness is defined by intentionality." [20]

Now I'm not claiming that Searle is actually committing himself to this sort of explicit equation of consciousness and intentionality; however, what he does say about the link between consciousness and intentionality suggests strongly that although the equation may not be there, the intimacy surely is. In his article "What is an Intentional State?", he's fairly explicit about this:

> What I actually believe to be the case... is something like the following: only beings capable of conscious states are capable of Intentional states.... And though any given Intentional state, such as a belief or a fear, may never be brought to consciousness, it is always in principle possible for the agent to bring his Intentional states to consciousness. [21]

Thus Searle is clearly holding that not only consciousness, but also conscious access to one's intentional states is prerequisite to intentionality. Notice that this connection is, once seen, apparent thoughout much of what he says in the "Minds, Brains, and Programs" responses: e.g., "I could have made the argument about pains, tickles, and anxiety...." [22]; "To interpret the symbol he would have to have some awareness of the causal relation...." [23]

Indeed, many points made in the course of this discussion in terms of "knowing that" or "understanding that" become much clearer when it is kept in mind that for Searle, these sorts of relationships are going to presuppose potential, if not realized, "consciousness of".

We can now see what at least part of the "causal powers" had by the brain (in virtue of its internal operations) necessary to its ability to secrete the juice of meaning are: the power to produce conscious mental states which have the right sort of relationship to the semantic properties of the intentional states of the brain (or organism). Now as intrinsic intentionality is, for Searle, the mark of the mental, it's not surprising that it's bound up with consciousness in an important way; intuitions about consciousness are, after all, central to anyone's pretheoretic notion of mind. However, as it's <u>semantic</u> properties which Searle thinks are essentially lacking without the right biochemistry, it's worth considering just how the semantic properties of the representational states of the brain -- the properties of <u>meaning</u> and being <u>about</u> certain things, of having particular referents and truth conditions -- are supposed to be intrinsically bound up with the internal operations of the brain, including the power it has to produce conscious mental states of the sort we have.

Now as Searle puts it at one point, the underlying idea here is something like the following:

> The brain is all we have for the purpose of representing the world to ourselves and everything we can use must be inside the brain. Each of our beliefs must be possible for a being who is a brain in a vat because each of us is precisely a brain in a vat; the vat is a skull and the 'messages' coming in are coming in by way of impacts on the nervous system. [24]

But one question which needs to be asked is precisely the one which is begged here. From the assumption that mental processes must use only "internal" properties of the brain, does it follow that the representational states of the brain cannot have semantic properties which aren't reflected in the biochemical properties of the representations? Can it make a difference what the signals are coming from? And for those which are so reflected, is it clear that they are not also reflected in the computational properties of the representations? I'll now turn to a discussion of Searle's position on semantics with an eye toward answering these questions.

III. Semantics (1): Indexicals

Now on one straightforward reading of what Searle has to say about the sort of contribution his "intrinsic intentionality" makes in the fixation of the "aboutness" relations of mental states such as meaning, reference,a and truth conditions, he would appear to be holding an absurdly strong version of the (these days somewhat discredited) view that meaning is "in the head". For he does claim that (1) "[mental representations] are defined in terms of their

content" [25]; (2) "the [intentional] object of a mental [representation] is just the actual object or state of affairs represented by [the relevantly related] intentional state" [26]; (3) "any [mental] representation is internally related to its [intentional] object in the sense that it could not be <u>that</u> representation if it did not have THAT object" [27]; and (4) "it is the operation of the brain and not the impact of the outside world that matters for the contents of our intentional states." [28] On what looks like the most natural reading of all this, the view presented would seem to be vulnerable to the following obvious sort of counterexample: Surely <u>everything</u> in my head might be the same on each of two occasions where I think "That man is a spy", but the situations differ in that on one occasion I was looking and pointing at Ralph (and thus referring to <u>him</u>), and on the other I was looking and pointing at Sam, Ralph's identical twin brother (and thus referring to <u>him</u>). To whom I've referred depends on who's actually there. But we're then, I take it, quite inclined to say that if these thoughts have states of affairs as intentional objects, the former thought's intentional object is the state of affairs consisting in <u>Ralph's</u> being a spy, and the latter thought's intentional object is the state of affairs consisting in <u>Sam's</u> being a spy. Hence, on the reading at hand, what's in the head (at least in Searle's sense) doesn't fully determine the intentional objects of mental states,

and thus can't fully determine the content of such states either -- directly contradicting (4).

I think that the misleading claim among the four cited is (3). In his article "Intentionality and the Use of Language" (from which claim (3) was taken), Searle seems to significantly weaken this thesis (at least implicitly) soon after advancing it. Indeed, on the very next page, immediately after making claim (2), he goes on to say that "if there is no such actual object or state of affairs represented then the intentional state does not have an intentional object though it does still contain a representation." [29] Given the intimate sort of relationship between content and object required by thesis (3), one is immediately prompted into wondering what state is "the" state Searle is talking about; if having the same intentional object is required in order to be the same intentional state, there just can't be any one state which may or may not have an intentional object.

Searle's position is not quite this easy to defeat, however. The trick here (and the position Searle clearly intends) is to read (3) as being about <u>tokens</u> of intentional states rather than types. Each token mental state here is taken to be in a certain sense "self-referential", and it is this which allows the type identical mental states to have different intentional objects -- indeed, to have different contents. This move is made most clearly in Searle's book, <u>Intentionality: An</u>

Essay in the Philosophy of Mind. Here, in considering an example involving the intentionality of visual perception, the point is put like this:

> ...type-identical visual experiences can have different conditions of satisfaction and therefore different Intentional contents. Two "phenomenologically" identical experiences can have different contents because each experience is self-referential. Thus, for example, suppose two identical twins have type-identical visual experiences while looking at two different but type-identical station wagons at the same time in type-identical lighting conditions and surrounding contexts. Still, the conditions of satisfaction can be different. Twin number one requires a station wagon causing his visual experience and twin number two requires a station wagon causing his numerically different visual experience. Same phenomenology; different contents and therefore different conditions of satisfaction.... [30] The conditions of satisfaction are: that there is a yellow station wagon in front of X and the fact that there is a yellow station wagon in front of X is causing the visual experience. [31]

Given this, the treatment of the earlier example is fairly straightforward. Although I have type-identical mental states in the two cases, the contents of the tokens differ in that each makes explicit direct reference to itself -- a particular token mental state. In each case, the conditions of satisfaction might be stated roughly as "there is a man over there causing this (token) experience and ...", but the difference in reference of "this (token) experience" allows the two to have different conditions of satisfaction and thus different intentional objects.

However, we're not out of the woods yet. Even if the general idea of this sort of analysis is accepted, there is

a clear way in which representational content in the present sense doesn't fully determine intentional object. What is actually referred to still depends on the external context of the token mental state. If I'm hallucinating, then my thought has no intentional object; if Sam's there, then it's him; and if Ralph's there, it's him. There's at least some inclination to say something like this: Pick one of the two token thoughts involved in our puzzle. Now why shouldn't we say that that very token doesn't fully determine an intentional object? After all, in different possible external contexts that token would pick out different objects.

What's misleading here is the reference to the intentional object as <u>that</u> object. To avoid the present problems, thesis (3) clearly must be taken as concerning the intentional object as given via a particular description and not as simply concerning that very object. Of course my token state could be that very one even if it didn't have the man who is in fact causing the experience (i.e. Ralph) as its object, but it must have as its object whatever man happens to cause the experience. The necessity of (3) is, to put it in a way Searle doesn't like, <u>de dicto</u> with respect to a characterization of the conditions of satisfaction rather than (<u>de re</u>) concerning the object which in fact satisfies such conditions. The object of the state in question must be the man (if there is one) standing in front of me causing that very experience; but it needn't

be Ralph, even though he is in fact that man.

However, if this is the way (3) is to be handled, it's not only stated a bit misleadingly, but it's also somewhat vapid. Of course, given a specification of the conditions of satisfaction of an intentional state, a token couldn't even be a state of that type unless it had as object whatever (if anything) meets those conditions. It's not, on this approach, the representational content alone which makes one thought about Ralph and the other about Sam; it's also the fact that the external world is set up in such a way that it's Ralph who happens to be the man who is in fact causing that experience. It would then seem that, at least in this kind of case involving explicit indexicals, content will only fix reference given an external context; and so at least this aspect of "aboutness" is not captured by "representational content" in Searle's sense.

But surely fixing referents for indexicals is problematic on most anyone's account; and it's been a standard move in philsophy of language to separate "intension" (in the sense of something like cognitive significance) from fixation of reference for such terms. As long as we have in hand a semantic characterization of the intentional state which fixes the referent, and some reasonable characterization of what external factors are relevant to the reference of indexicals (like what's around at the moment), things don't seem so bad. Indeed, any semantic account of a "psychological states in the narrow

sense" (which the cognitivist would seem to want as well) must find a way to deal with the problem of explicit indexicals. What either sort of view needs in answering questions as to what indexicals are "about" is to give a kind of relativization to a context of something like the de re reading of thesis (3). However, what I hope to show now is that for Searle, the kind of problem found here is going to turn out to be contagious in such a way that far more than the explicit indexicals are affected.

IV. Semantics (2): Non-indexicals

Let me illustrate this by means of a variant on the "inverted spectrum" case. I think that the sort of "inverted intentionality" situation created for Searle here goes far beyond this sort of spectrum inversion case -- I'll say a bit more about this later. However, I think that this particular sort of example provides a good illustration of the problems involved here.

Consider the following crazy sort of case. [32] A few centuries down the road, we stumble upon a planet which has the peculiar property of having its colorations reversed; or to make things simple, just green and red: the planet looks just like Earth in every way you like except that "grass" is red, "roses" are green, and so on. Furthermore, the inhabitants of this planet have bodies and brains just like ours, with only the following exception: red/green color inverting lenses naturally cover their eyes, so that the

color inversions in their world are, to put it a bit misleadingly, righted -- when an alien looks at his (green) roses, he has exactly the same type of physical events occur in his brain as would some doppelganger of his on Earth upon looking at his (red) roses. Of course, to get it so that everything in the brains of a pair of doppelgangers is the same in such a situation, we could suppose that some of our aliens speak a language which is just like English, except that in their language, the word 'red' is associated with the color of their roses and apples (which are, recall, green), and 'green' is associated with the color of their grass and trees (which are red). [33]

Now consider such a pair of doppelgangers; call the Earthling 'Bob E." and the alien 'Bob A.". Now if Bob E. looks at something red and thinks to himself "That's red", then the conditions of satisfaction of his intentional state involve their being a red object in the world causing his visual experience; if his thought is true, there must be a red object so situated. But if Bob A. were to have exactly the same things occur in his brain (and thus have a conscious experience with, as we might put it, the same phenomenological character as Bob E.'s), surely the conditions of satisfaction of his intentional state involve there being a green object in the world which is the cause of his visual experience. Green objects are just the ones Bob A. (and the other aliens) always pick out by

the word 'red' and the phenomenological character linked with it.

What the inverting lenses, inverted world coloration, and language alteration have done is to manipuate the head/world relationship so that thoughts which are instantiated by type identical brain states and have the same phenomenological character are about (or "directed upon") red when thought by an Earthling, and about green when thought by an alien. 'Red' in the alien's language refers to -- indeed, means -- the same thing that 'green' does in English (and vice versa), in spite of the fact that the brain states and phenomenological characters linked with each word are the same in the two cases. So, we again have a case where type identity of brain state is compatible with difference in conditions of satisfaction of intentional state; and rather than involving explicit idexicals, for which cognitive significance and the fixation of reference have long been seen as requiring separation, it involves color terms -- terms used to pick out what are paradigmatic examples of "secondary" properties.

One very bad option for responding to this would be simply to reject the central claim of the foregoing argument; i.e., to deny that the aliens actually mean and refer to green by their use of 'red', and that the state Bob A. has which has exactly the same phenomenological character as that caused in Bob E. by his seeing a red object is

actually about or directed upon a green object (even though those are the kind that typically cause that state in him). I take it that there is something patently absurd in such a line. The aliens satisfy any sort of criterion of use you like for referring to green things by their use of 'red' (e.g., consistently pointing at green things and saying "That's red"). In any straightforward sense, it's clear that they have learned to use 'red' to pick out green things, and use 'red' in just the way we use 'green'; it would be small consolation to little alien Johnny, blurting out between sobs, "But I wanted a red one, not a green one!" to be told that he was mistaken, and that he had gotten what he wanted after all. And of course, the situation is entirely symmetrical: the alien Searle would have all the same sorts of justification for claiming that it's those Earthlings who've gotten it wrong; Earthling Johnny wuld fare no better there than would his doppelganger here. Surely for us or for the aliens, such an account reeks of the wildest sort of chauvinism.

This sort of move would seem to epitomize exactly what's wrong with what Putnam has called "magical" theories of reference -- accounts which hold that the reference relation obtains purely in virtue of some wholly unexplainable (hence "magical") connection between certain sorts of representations (on Searle's story, the intrinsically intentional ones) and the objects of those representations, totally independently of any kind of causal

or contextual link. [34] Problems involving such "magical" theories of reference and the consequences of rejecting them are extremely interesting, but (I hope) a bit outside the scope of the present paper. However, let me note just one point here which is particularly relevant to the present discussion: On such a "magical" line, it apparently could be the case that we're the ones who have gotten it wrong, and that, say, the phemomenological character which is typically caused in me by my looking at red things is actually "magically" linked up with green things instead. But surely it seems as though this is the sort of thing that we can't have gotten wrong -- at least not because we turn out to be (somehow) like our aliens. To commit ones' to this much is, furthermore, not to buy onto any sort of radical anti-realism; it's not to hold that we couldn't end up "getting things wrong" in some way regardless of our sceintific successes, but just to hold that the particular sort of difference between the aliens and ourselves isn't the kind which could affect what our words and thoughts refer to and are true and false of.

So I take it that a line like this is out, and that it must be acknowledged that our aliens do refer to green by their use of the word 'red' and the phemomenological character linked with that word (in them as well as in us). We already have one kind of case for which phenomenology doesn't fully determine conditions of satisfaction, and Searle is perfectly prepared to live with this; as he

admits, "a man and his Doppelganger can be type-identical down to the last microparticle, and their Intentional contents can still be different; they can have different conditions of satisfaction." [35] The problem is then first of all to show how to subsume this new case under the "self-referential" move.

How such a move is to go is fairly obvious, The general idea is of course to hold that the conditions of satisfaction of Bob E.'s thinking "That's red" involve there being the sort of thing which typically looks red to him causing his visual experience; and the conditions of satisfaction of Bob A.'s phenomenologically identical thought involve there being the sort of thing which typically looks red to him (i.e. a green thing) causing his visual experience. Indeed, a move very much like this is offered by Searle in an attempt to answer Putnam's "Twin-Earth" case from "The Meaning of Meaning", in which doppelgangers of ours on "Twin-Earth" refer to something different tha we do by 'water' because the stuff which plays the role of water (fills lakes, good for drinking, etc.) is actually a different chemical substance. Searle's line:

> The indexical definition given by Jones on earth of 'water' can be analyzed as follows: 'water' is defined indexically as whatever is identical in structure with the stuff causing this visual experience, whatever that structure is. And the analysis for twin Jones on twin earth is: 'water' is defined indexically as whatever is identical in structure with the stuff causing this visual experience whatever that structure is. Thus, in

> each case we have type-identical experiences, type-identical utterances, but in fact something different is meant. That is, in each case the conditions of satisfaction established by the mental content (in the head) is different because of the causal self-referentiality of perceptual experiences. [36]

Now I think that there are deep and special puzzles involving the meanings of natural-kind terms, and I'm going to avoid them here. [37] But at least this much is obviously wrong with the portion of Searle's account given here: Surely the meaning of 'water' isn't just "same kind of stuff as is causing this visual experience". Hold some rubbing alcohol in front of me and tell me it's water and I may believe it; but that doesn't make 'water' in my idiolect refer to alcohol. But Searle really doesn't think it does; the quoted passage is just a bit misleading in this respect. What he does think about this comes out more clearly in his discussion of what he calls "the problem of particularity."

Here, in order to avoid a problem which is like this water/alcohol one -- only stated in terms of reference to a particular person rather than a particular kind of stuff -- he suggests that the conditions of satisfaction in such cases involve in part connection to the objects of past experiences. In an example where he seeks to bring out how the conditions of satisfaction of Jones' thought that he's seeing Sally require that it's <u>Sally</u> that he's seeing and not her double, he makes this explicit. I'll avoid the details of the discussion, but the general idea ought to be fairly clear; as Searle at one point puts it, a "way of

describing the situation pretheoretically might be, 'I am now seeing the woman I have always known as Sally'." [38] But for present purposes, it's critical to see what such an identification is acknowledged to depend on:

> ...in order that it be part of the conditions of satisfaction of Jones' Intentional state it must be caused by Sally rather than twin Sally, Jones must have some prior identification of Sally as Sally, and his present experience must make reference to that prior identification in the determination of the causal conditions of satisfaction. [39]

All this in hand, let's consider the situation of the spectrum inversion case. Surely the general move must be the same, as suggested earlier -- both our Earthling and alien will have conditions of satisfaction for 'red' and the associated experiences which will be roughly "what I have always known as red". The difference comes in through the different reference of "I". However, we're in the same situation now with regard to "red" as we were with the indexicals: What fixes what "red" is about is not the intentional content alone; a given state (even a given self-referential token) depends for it's semantic properties on the way the external world is -- on what happened to cause it. And in this case, it's even worse. First of all, the intuition seems strong that, although it was only the reference and not the lexical meaning of the indexicals which varied in different contexts, the meaning of 'red' is different for the aliens and Earthlings, even though their phenomenological characters are the same. Second,

this move leaves us with a characterization of meaning which gives us very little of what a story about meaning should -- in particular, it has a fundamental difficulty with interpersonal sameness of meaning. And third, the kind of problem presented here -- and the possibility of generalizing it -- suggests that "intrinsic inner content" doesn't give a means for isolating brains from computers in the way Searle wants after all.

On the first point, I really don't have much to say, except to suggest that any sort of non-question-begging characterization of meanings -- such as by the role they play in the explanation of behavior -- would seem to support this suggestion. The earlier case of the two Johnnys and "wanting a red one" seems to be exactly the kind of example which supports this. It's worth pointing out here, though, that it's this sort of role which separates the current case from the standard "twin earth" case using natural kind terms like 'water'. I for one am inclined to accept Putanam's intuitions about 'water': in a world in which the stuff around which looks, tastes, etc. just like water is really not H20, 'water' not only refers to something different, but in a certain sense of meaning, means something different as well. However, given that we explain behavior in terms of the content of mental states, and the behavior of my doppelganger (in Putnam's case) and myself is identical, there's at least some plausibility to the idea that there is a kind of meaning -- "narrow content" -- which

we do share. But in the spectrum inversion case, meaning as it figures in the explanation of behavior is varied within the bounds of sameness of brain state. So even if (as I'm inclined to think) some such notion of "narrow content" can be made clear, and the problems with indexicals and natural-kind terms can be dealt with, Searle's line still faces the present problem.

The second point is this: Surely one of the things an account of meaning ought to least suggest is something about what it is for two people to mean the same thing by a word, representation, or whatever. But on Searle's story, not only do we not get this, we get the result that two people can't mean the same thing. If what I mean by 'red' involves direct reference to me and my experiences, and what you mean by 'red' involves direct reference to you and your experiences, then we don't mean the same thing by 'red'. The spectrum inversion case shows that thinking of something as "looks red to me" won't give a criterion of sameness of meaning because it gives the wrong answer. To get any account of sameness of meaning, then, it looks as though you're going to have to go outside of the brain, and into the world -- at least to the level of proximal stimuli, and maybe further.

The third point is, for present purposes, critical. Recall what all of this "intrinsic inner content" talk was in support of: the view that the states of brains have their intentional properties (a) purely in virtue of their

internal operation, and (b) in virtue of facts about their internal operation other than those concerning their information-processing structure. What we have seen so far is that what's in the brain doesn't determine the semantic properties of our mental states as completely as Searle would seem to suggest. Phenomenological character and biochemical structure can be held constant through significant changes in meaning; and there isn't anything about the brain in a vat that makes its "phenomenologically red" thoughts mean red rather than green. But if the meanings of intentional states can in this way be varied even within the bounds of sameness of brain state type, we are then left with the question we started with: What's the relationship between Searle's purely internal content and the contents of our intentional states?

It's worth noting here that it doesn't look as though this problem is limitied to the particular example used. It's obviously the type of case which is of concern -- the type in which brains have both different sensory apparatus and different external environments, and these happen to mesh in such a way that the input/output relations of the brains (and so their inner states) stay the same, while at the same time the difference in how the brains are related to the world alters what their internal states are about. Now I'm inclined to think that this sort of permutation -- within the bounds of sameness of brain state type -- of what we might call external conditions of

satisfaction can be made for many purely internally individuated representations; however, I won't argue the point any further here. [40] It's worth pointing out here, though, that it's the cases where meaning seems to depend to some extent on "internal" connections between the meanings of representations (unlike the color case) for which this permutation becomes less plausible -- but it's just this internal complexity which makes more plausible the view that it's something like inferential structure placing constraints on this permutation. And the question is, given that meaning can be permuted within the bounds of sameness of biochemistry, why should we think that the constraints on such permutation are not just those set by computational structure?

Let me draw at least this much of a moral from all this discussion: Whatever Searle's purely inner notion of content comes to, it has a whole lot less to do with telling us what given mental states mean than he would appear to be suggesting. It's not that I think there isn't any way to make sense out of the notion of what the internal contribution to content is; indeed, part of spelling out the cognitivistic account of the mental depends on being able to do something like this. But this much seem clear: the notions of meaning and content which are around are such that a good deal of meaning depends on relationships to the external environment -- at least to proximal stimuli. And in this sense, we are like the computer in which the

same program has two different interpretations on two different days. As far as my "internal life" or phenomenology goes, it doesn't matter whether I'm an Earthling or an alien; but the meanings of my words and thoughts differ, depending on which I am.

V. Robots and the "Empirical" Question

Now if this is at all right, it's clear that in considering whether a given machine's operation determines a semantic interpretation for its states in the sort of way our brain's operation does, we should be concerned with whether it gets the same sort of semantic interpretation given the same relation to the world (as much as possible) that we have. That is, we should consider whether a robot which is computationally equivalent to one of us, and is such that its computer "brain" is hooked up to a body so as to enable it to go about the world in much the way you and I do, would then have representational states which refer to the world around it the way ours do. Now Searle in fact considers just such a case, and has the following to say about it:

> I entirely agree that in such a case we would find it rational and indeed irresistable to accept the hypothesis that the robot indeed had intentionality, as long as we knew nothing more about it.... but as soon as we knew the behavior was the result of a formal program, and the actual causal properties of the physical substance were irrelevant, we would then abandon the assumption of intentionality. [41]

Now on one way of viewing what is said here, it's just absurd. Of course it couldn't turn out that the actual causal properties of the physical substance making up the computer "brain" were totally irrelevant. The computer couldn't cause the robot to make movements and noises at all unless there were certain sorts of causal powers had by at least some of its physical components; in particular, those powers by means of which it can pull levers, trip relays, fire neurons, or whatever it has to do in order to get the body to respond to its instructions. Furthermore, the non-computational physical properties of the machine must be in a certain sense relevant given the constrants of time and space: computers made out of certain sorts of materials just wouldn't be able to instantiate a program of anything like the complexity which must be involved while (a) fitting inside a medium-sized head (or body), or (b) running the program in "real time" -- fast enough to allow the robot to interact with the environment like we do. Even if one allowed for radio links or some such thing so that the computer didn't have to fit in the body, I take it it's clear that, say, an "homunculi-head" with real human beings for homunculi just wouldn't be able to push the symbols around fast enough. As Fodor puts it in his response to Searle, "it might be, in point of physical fact, that only things that have the same simultaneous weight, density, and shade of gray that brains have can do the things that brains can. This would be surprising, but it's hard to see why a

psychologist should care much." [42]

Surely Searle's intent is not then to suggest that the physical properties of the computer might be entirely irrelevant, but rather that they might be irrelevant with respect to the content of the computer's states; that is, the claim is that the physical substance might not be the sort which oozes intentionality, in which case the particular properties of the physical substance would not be contributing any content to the computer's states. Given Searle's line on biochemistry as the source of meaning, something like this would seem to be the natural reading of the above passage. However, even if we take the "irrelevance" of the "actual causal properties of the substance" in this way, the claim still seems to be inherently puzzling. For there's at least some substantial inclination to read Searle as claiming here that finding out that the robot's behavior was "the result of a formal program" would be itself sufficient grounds for rejecting the assumption of mentality, and that it's not required that we somehow make an additional discovery that "the actual causal properties of the substance were irrelevant." Indeed, he at one point seems to put this fairly explicitly, by claiming that "If we knew independently how to account for its behavior without such assumptions [i.e. of mentality] we would not attribute intentionality to it." [43]

But then given the assumption of our own

intentionality, it can't turn out that our own behavior is (or could be accounted for as) "the result of a formal program" -- which is to say, I take it, that there cannot turn out to be any description of our brains as automated formal systems, the instantiation of which is causally sufficient (given the right hookup with the body and the right dasein) for the production of the sorts of behavior which we in fact produce. But what then are we to say about the behavior of our robot, which is cognitively equivalent to me and has the right bodily hookups and position in the world? Searle surely continues to refer to the behavior of such a robot even after the "assumption of intentionality" has been rejected. What sorts of behavior does it produce, if not just the same sorts that I do?

Searle's answer here, I take it, would be just what one would expect in the light of the earlier discussion regarding the distinction between intrinsic intentionality and observer-relative ascriptions of intentionality. In another context, while discussing the behavior of performing speech acts, Searle has the following to say:

> To characterize [states] as beliefs, fears, hopes and desires is already to ascribe intentionality to them. But speech acts have a physical level of realization, qua speech acts, that is not intrinsically intentional. There is nothing intrinsically intentional about the utterance act, that is, the noises that come out of my mouth or the marks that I make on the page. [44]

The result of carrying this sort of view across to the

present discussion is fairly clear. (For simplicity, let me just talk about that subclass our behavior consisting of our utterances; surely this is the most interesting part of our behavior when our worry is about content, and I think nothing critical is lost for the particular point at hand.) Searle is, I think, perfectly willing to allow that a robot which is cognitively equivalent to me produces the same utterances as I do, considered as acoustic waveforms, movements of articulatory apparatus, strings of phonemes, or perhaps even syntactic forms. What he won't allow is that being such a robot could be itself sufficient for the production of the same sorts of utterances as I produce considered as speech acts which are the expressions of certain contents.

There's something right in this line, and it's what Searle tries to get at in another place in claiming that "rules affecting human behavior... are defined by their content, not their form." [45] When the cognitive model is considered as a purely syntactic machine of a sort, I'm inclined to agrree with Searle here: to give an account of the noises we make in terms of form (this case, something like how they come about as a result of syntactically characterized computational activity and its interconnection with our articulatory apparatus) is not to give an account of them as meaningful bits of human behavior. Surely this point is well taken when these noises are accounted for in terms of the neural (or chemical, or

microphysical) structures which produce them. To give a mentalistic account of the etiology of these noises is to consider them as contentful utterances, and then to explain how an utterance with that content came to be made; and to give such an account, it's not sufficient to explain physically how a certain accoustic waveform came to be made, and then simply point out that the waveform has the same phonemic structure as do the English words "it's raining".

Given all this, what cognitive science must do in order to distinguish itself from these sorts of explanations of, say, the noises we make, is to show how it gives an account of how, say, that speech act -- considered as an act of expressing some specific content -- was produced. Now what seems to be the standard story of how this might be done is this: [46] Cognitive science is to account for the production of behaviors individuated by content by showing how they are the result of the subject's being an instantiation of a certain sort of semantically interpreted computational / representational system. So the account given of why, for example, Sam asserted that it's raining will involve showing how standing in certain sorts of computational relations to semantically interpreted formulas -- e.g. his standing in the computational correlate of belief to formulas which have the interpretations "it's raining" and "I've just been asked what the weather is like" and so on -- comes to cause his utterance, the meaning

of which is that it's raining.

Now I take it that Searle is not trying to give any sort of a priori argument that there can be no such cognitive discription and theory of the way we process information; surely this is the sort of question that should be decided on the basis of future empirical successes and failures in cognitive psychology. Rather, his claim is that the semantic interpretation of such representational states must come from the physical or chemical character of the physical realizations of any such representational systems. But note that any such account of the production of our behavior as the result of the workings of some semantically interpreted computational system will give us just as good an account of the behavior of our robot as it will of the behavior of a person. We would, by bringing such an account to bear on our robot, be explaining its actions in terms of their contents. Of course on Searle's line, we would be giving an account of its behavior via some wild indulgence in observer-relative ascriptions of intentionality; as its "brain" lacks the right stuff for intrinsically intentional representations. Given this, we can make sense out of Searle's original assertion regarding the robot: If the robot -- or one of us -- has intrinsically intentional states, it can't be the case that our behavior is the result of a formal program in the sense that it has whatever content it does in virtue of being prduced by that program (hooked up to a body and situated in the world in

such-and-such a way). On this story, the robot's behavior has all the content it does (which is all observer-relative) in virtue of these sorts of consideration, whereas our behavior has content (intrinsically) in virtue of some additional facts.

Of course the problem now is that it's extremely difficult to see how it is that, as Searle says repeatedly, whether and entity has intrinsically intentional states or not is an empirical question; e.g.:

> ...perhaps, for example, Martians also have intentionality but their brains are made of different stuff. That is an empirical question, rather like the question of whether photosynthesis can be done by something with a chemistry different from that of chlorophyll.... indeed it might be possible to produce consciousness, intentionality, and all the rest of it using some other sorts of chemical priciples than those that human beings use. It is, as I said, an empirical question. [47]

Now as I noted earlier, it's of course an empirical question as to what other kinds of physical stuff could be made into a computer which could instantiate the right program at the right speed in order to interact with the world in the way we do. For Searle, however, it's clear that given a robot which is cognitively equivalent to me, and which interacts with it's environment in the same way, it's still an empirical question whether or not that robot has intrinsic intentionality; and this is true even if (a) entities with intentionality (like us) are such that everything relevant that we do (including internal mental

activity) is, considered under the best account of "observer-relative" ascriptions of intentionality content we might come up with, a result of our having a given sort of cognitive structure and dasein; and (b) such "observer-relative" ascriptions of intentionality match up (in our case) with whatever the right sort of ascriptions of intrinsic intentionality are. (Notice the striking similarity between observer-relative ascriptions of intentionality and ascriptions of intrinsic content.) Anything I might say about the content of your states and actions, I could also say about the ("observer-relative") content of those of a cognitively equivalent robot. Surely we meet up here with the classic bugaboo: What possile reason is there for saying in the case of the robot that the ascriptions don't truly ascribe intrinsic intentionality? What possible empirical test could tell us whether or not our latest creation managed to have l'etre-pour-soi, or whether God has seen fit to spit a little drop of ectoplasm into its head? How could we tell unless we could get inside its head and see what (if anything) it's like to be it?

Indeed, in his response to Dennett's commentary on the "Minds, Brains, and Programs" piece, Searle does make the suggestion (in connection with the discussion of one of his "Chinese room" variants involving the homunculus memorizing the rules of the appropriate program) that an example of this kind given by Dennett "is underdescribed, because we

are never told what is going on in the mind of the agent." [48] He goes on to offer the following admonishment: "Remember, in these discussions, always insist on the first person point of view. The first step in the operationalist sleight of hand occurs when we try to figure out how we would know what it would be like for others." [49] Now surely it's not that Searle thinks that, in general, if one were (or were to ask) the homunculus in a given machine, one would know (or be told) whether or not the flame of consciouness is present and related to the states of the entity in question in the right way to make the representations being manipulated intrinsically intentional ones. In the first place, this would conflict with Searl 's line on "Haugeland's demon", a speedy little homunculus who zips around tickling neurons -- in a brain in which the neurons have been chemically isolated from one another -- in just the way they would have been tickled had they not been so isolated. Intentionality is produced in this way (or so Searle says); but surely the demon needn't know this. Secondly, such a criterion would seem to have its applicability limited to single-homunculus based machines; and given the notion of a program at hand (which is, recall, substantially stronger than that of simply computing the same function), it's far from obvious that, say, my cognitive program could be instantiated on a single-homunculus based machine at all.

Rather, then, it would seem that what Searle is playing

on here is the suggestion that in the case of some person internalizing the program, the internalizing and the (alleged) internalized mind must somehow be the same; that is, the real suggestion is that if one were the machine itself, the embodiment of the program, then one would know whether or not intentionality resided there. Of course, to make Nagel's distinction [50], the claim shouldn't be taken as "if I were the machine..."; for if I were it, there would surely be something it would be like, simply because there's somthing it's like to be me, whatever I'm up to. Instead, the question is whether there's anything it's like for the machine to be it. But if this is what Searle's empirical claim is about, I'm totally puzzled. I'm quite inclined to think that the question of whether or not there's something it's like to be somthing does a very good job of capturing a fundamental intuition about consciousness; but I just don't see any way to milk a "testable empirical criterion" out of it. Surely the burden of proof is on those who might wish to hold otherwise.

In any case, it's clear where to classify Searle's worries about intrinsic intenionality among the two sorts of strategies for arguing against cognitive science which John Haugeland distinguishes in his article "Semantic Engines: An Introduction to Mind Design":

> The first, or hollow shell strategy has the following form: no matter how well a (mere) semantic engine acts as if it understands, etc., it can't really understand anything, because it isn't (or hasn't got) "X" (for some

> "X").... The other, or poor substitute strategy draws the line sooner: it denies that (mere) semantic ingines are capable even of acting as if they understood -- semantic engine robots are not going to get that good in the first place. [51]

Surely Searle's argument is paridigmatically of the first sort; and with regard to this strategy, Haugeland goes on to list what he sees as three leading candidates for "X": Consciousness, primary (or intrinsic) intentionality, and caring. I hope the moral of the preceding discussion is clear: There may in fact be deep and independent worries about the second candidate, but the ones which Searle gives us seem to be purely derivative from worries about the first. Now if we could just clear those up....

NOTES

[1] Leibniz (1961), p. 206.

[2] Searle (1980).

[3] Searle (1980), p. 418.

[4] See Block (1981a).

[5] I take it that neurophysiologists don't really care much whether the "important" properties are functional or physical in the present sense.

[6] See Block (1981b).

[7] Searle (1980), p. 422.

[8] Lycan (1980), p. 435.

[9] Meaningless bit of fluff like this, and somebody tries to make something out of it. Dreyfus (1980) distinguishes between what he calls "Dasein1", "which is something like man's actual embedding in the physical universe," and "Dasein2", the "background of already entrenched social practices" -- which are in a certain sense internalized -- against which "our activity of taking-to-refer and claiming-to-be-true takes place." It's Dasein2 which Dreyfus thinks is actually "being-in-the-world" in the Heideggerian sense, and he thinks it's this which presents problems for cognitivism.

Sticking to the present issue, however, it's clear that it's Daseinl that's involved at present, although Dreyfus' gloss on this might be a little misleading. There's no prima facie reason that all sorts of facts about our external surroundings (including social ones) might not be relevant parts of one's Daseinl, in the sense the they enter into the fixation of reference for our mental states.

[10] Searle (1980), p. 452.

[11] Ibid.

[12] Ibid.

[13] This sort of example is suggested in Putnam (1981), chapter 1. Note, however, that I'm not talking about the sort of "full-blown" brains-in-a-vat case that Putnam is worrying us with, in which all the sentient beings in the universe are brains in a vat. The sorts of problems involved in such a case are extremely interesting, but beyond the scope of the present discussion. Rather, for the present case it would seem that what we want is to hold our own situation fixed (as not being brains in a vat) and then consider a brain in a vat existing in our world.

[14] All the hedging here is because even this sort of causal "divorcing" of the brain from the rest of the world might not be enough to guarantee that the content of its states was somehow purely a result of it's biochemical makeup -- there still might be the possibility of some sort

of counterfactual story about the content its states would have had, had things been different. As I hope will become clear later on, the issue here about counterfactuals and how they are to be constrained is in a sense the whole ball game.

[15] Rey (1980), p. 91.

[16] Searle (1980), pp. 451-2.

[17] Searle (1980), p. 420.

[18] Searle (1980), p. 452.

[19] Husserl (1962), pp. 228, 231.

[20] Sartre (1957), p. 38.

[21] Searle (1979a), p. 92.

[22] Searle (1980), p, 453.

[23] Searle [1980], p. 454.

[24] Searle [1983], p. 230.

[25] Searle (1980), p. 423.

[26] Searle (1979b), p. 185.

[27] Searle (1979b), p. 184.

[28] Searle (1980), p. 452.

[29] Searle (1979b), p. 185.

[30] Searle (1983), p. 50.

[31] Searle (1983), p. 61.

[32] A similar case is discussed in Harman (forthcoming).

[33] This is probably overkill of a sort, as I really don't think anybody would want to hang anything on the difference between the brain states of a pair of doppelgangers which was simply a result of their having learned to link up different words with different mental states (or however this difference should be described). Nonetheless, overkill or not, we've now set things up so that we can get perfect type identity of brain states in an alien / Earthling pair as they ponder the colors of their respective roses, or of their respective lawns.

[34] See Putnam (1981), especially chapters 1 and 2.

[35] Searle (1983), p. 207.

[36] Searle (1983), pp. 207-8.

[37] These are discussed in part 2 of this thesis.

[38] Searle (1983), p. 68.

[39] Searle (1983), p. 66.

[40] Some of the possibilities are discussed in Putnam (1981), especially chapter 2.

[41] Searle (1980), p. 421.

[42] Fodor (1980b), p. 432.

[43] Searle (1980), p. 421.

[44] Searle (1979a), pp. 88-9.

[45] Searle (1980), p. 454.

[46] See Fodor (1980a).

[47] Searle (1980), p. 422.

[48] Searle (1980), p. 451.

[49] Ibid.

[50] See Nagel (1974).

[51] Haugeland (1981), p. 32

Part 2:

MEANING PSYCHOLOGIZED

I. Introduction

The by now familiar story goes like this: if cognitive science is to give a reconstruction of the pretheoretical notions of the mental -- belief / desire psychology, the characterization of mental states and representations in terms of their content, and so on -- then one needs a notion of the content of mental representations. Indeed, even if one's hopes for the cognitivist strategy are somewhat more modest, it looks as though one will need such a notion in understanding the nature of one of the constructs central to contemporary cognitive psychology -- that of semantic storage. But over the past few years, a number of philosophers have put forward arguments purporting to show that just such a notion is fundamentally problematic. In what follows, I'll consider some of these philosophical worries about the semantic properties of mental representations, and suggest what about them, if anything, should concern those interested in the current enterprise of cognitive science.

The central theme of these problems is just this: cognitive state, or cognitive significance of a representation, doesn't seem to determine what we normally take to be meaning. Now before I consider the question

of meaning explicitly, let me say a bit about the notion of cognitive state that's operative here. The central point to bring out is that this notion should be what has been called an "autonomous" or "solipsistic" one; that is, type identity of cognitive state ought to be guaranteed at least by physical type identity of subjects. More simply: cognitive state ought to be a characterization of our psychological subject itself, and should be in some sense indifferent to what goes on outside the subject. The point is one found throughout the literature, and I won't spend a lot of time arguing for it explictly here, instead choosing to focus on the problems involved in accepting it. But the central theme of such arguments is clear, and Stitch's "replacement argument" captures it as well as any: Surely our psychological theory ought to ignore differences which not only can't turn up in behavior, but which can't turn up in any characterization of the subject's internal structure (and so the way in which it produces that behavior) either; thus it ought to treat physical duplicates just the same -- i.e. physical type identity ought to entail psychological type identity.[1]

Now given both the idea that psychology ought to be autonomous or solipsistic, and the need to come up with a notion of the content of psychological states or mental respresentations, we must then specify some sort of content which at least supervenes on physical structure of the subject -- indeed, one might hope (given the character of

cognitive psychology), on something far less constraining, such as rough similarity of information processing structure. That is, we need a notion of what gets called narrow content. And in what follows, I'm going to assume something like this last point -- although I don't think it will generally make much difference. That is, I'll assume that whatever facts about intentional states and properties of a subject supervene on that particular physical structure would equally well supervene on anything which had the same cognitive structure, where this is taken to include not only a computational characterization of the subject's functional structure, but a "real time" characterization of the transducer states as well.[2] If you think that narrow content ought to supervene on less than this, it won't matter for the present purposes -- at least it should supervene on this.

One way to view the central problems posed for an accout of narrow content is to view them as stemming from the effects of two different sorts of context on the meaning of mental states; a natural characterization of these would be as external and internal context. The problem with the former (the standard "Twin Earth" problem) takes the form of the suggestion that even if we guarantee total physical type identity of subjects, changes in the external environments of our subjects have a critical effect on what we would normally want to say the content of the mental states of those subjects were; even though they

are internally identical, their words and thoughts clearly have different meanings, and thus meaning can't supervene on the individual alone. Thus (the story goes), intentional psychology can't be "individualistic".

The problem with the latter sort of context, which I find much more serious, is one which lies in wait for those who would find their way past the first problem. For once one does make sense of a notion of narrow content which is shared (at least) by doppelgangers, and which thus does not depend in the wrong way on facts about reference for individuation of meanings, then it is suggested that too much has been left behind for the notion at hand to be anything like the normal notion of meaning. The problem is that there will be no "coarse" enough way of individuating the content of mental states in cases of subjects whose internal structures aren't exactly the same without appeal to non-individualistic facts about what the symbols and words refer to. Or to put it slightly differently, the claim is that the only way in which we in fact are able to distinguish between <u>meanings</u> and other collateral information is by appeal to "non-autonomous" semantic considerations. In this paper, I'd like to deal in turn with each of these problems; offering a kind of solution to the first, and offering some sort of hope for dealing with the second -- the <u>real</u> "problem of narrow content".

II. External Context and Twin-Earth

Let me then turn to the first problem and to the standard "Twin Earth" case. The general idea of this sort of example is fairly clear, and is by now found scattered throughout the literature. Imagine a world which is just like Earth in every respect, down to having inhabitants which are microphysically type identical to the human inhabitants of Earth. Now, while keeping at least some of the inhabitants of "Twin Earth" absolute "internal replicas" of Earthlings, we imagine the external environment altered in different small ways and consider how these alterations affect ones intuitions about the semantic properties of the words and thoughts of the folk from Twin Earth. In this way, we can test against our intuitions the degree to which we might wish to say that "meaning is not in the head".

The device is originally Putnam's [3], and his first example is the best known: We are asked to imagine that on Twin Earth, the stuff which fills lakes and reseviors, is used for drinking and bathing, and which generally plays the role which water does on Earth, and which is on the whole more or less indistinguishable from water in its macro qualities (it's clear, oderless, tasteless, and so on), nonetheless has a chemical structure quite different from that of water -- rather than H2O, it has some structure which we can abbreviate XYZ. Putnam then asks the question: what does a Twin Earther refer to by "water"? And the answer which he gives is that he refers not to water (which is, of course, H2O), but rather to XYZ -- and

this in spite of the fact that exactly the same things go on inside the heads of Earthlings and their doppelgangers. Moral: psychological state (taken in the autonomous sense) doesn't determine extension; and thus in the sense of "meaning" in which meaning determines reference, "meanings ain't in the head."

Of course there is a class of cases for which its even clearer that meaning alone does not determine reference -- one for which no special science fiction stories need be told: the class of explicit indexicals. In "The Meaning of Meaning", Putnam notes the similarity of the cases using natural-kind terms (like "water") to these:

> Words like 'now', 'this', 'here', have long been recognized to be indexical, or token-reflexive -- i.e. to have an extension which varied from context to context or token to token. For these words no one has ever suggested the traditional theory that 'intension determines extension". To take our Twin Earth example: if I have a doppelganger on Twin Earth, then when I think 'I have a headache', he thinks 'I have a headache'. But the extension of the particular token 'I' in his verbalized thought is himself...while the extension of the token 'I' in my verbalized thought is me.... So the same word, 'I', has two different extensions in two different idiolects; but it does not follow that the concept I have of myself is in any way different from the concept my doppelganger has of himself.
>
> Now then, we have maintained that indexicality extends beyond the obviously indexical words and morphemes (e.g. the tenses of verbs). Our theory can be summarized as saying that words like 'water' have an unnoticed indexical component: 'water' is stuff that bears a certain similarity relation to the water around here. Water at another time or in another place or even in another possible world has to bear the relation same-L to our 'water' in order to

> be water. Thus the theory that (1) words have 'intensions', which are something like concepts associated with the words by speakers; and that (2) intension determines extension -- cannot be true of natural-kind terms like 'water' for the same reason the theory cannot be true of obviously indexical words like 'I'.[4]

Immediately following this, Putnam makes a point which he is later in the paper to reject (albiet somewhat weakly), but which certainly looks so far to be exactly right:

> The theory that natural-kind terms like 'water' are indexical leaves it open, however, whether to say that 'water' in the Twin Earth dialect of English has the same meaning as 'water' in the Earth dialect and a different extension (which is what we normally say about 'I' in different idiolects), thereby giving up the doctrine that 'meaning (intension) determines extension'; or to say, as we have chosen to do, that difference in extension is ipso facto a difference in meaning for natural-kind words, thereby giving up the doctrine that meanings are concepts, or, indeed, mental entities of any kind.[5]

Now before turning to a consideration of the reasons one might have for rejecting the former view here, let me digress a bit and say what such a view might look like in a little more detail. If we're going to assimilate the natural kind terms to the explicit indexicals, we had better have some sort of account of the semantics of the latter. A plausible start at such an account has been developed by David Kaplan, centrally in his manuscript "Demonstratives". Let me then breifly sketch this kind of an account, and say a bit about how natural-kind terms might be subsumed under it.[6]

III. Indexicals -- Content and Character

Kaplan's account is in effect a "two-tiered" story about the meaning of indexicals: he calls the two kinds of meaning "content" and "character". The character of an expression (indexical or not) is to be thought of as something like what Putnam calls the concept associated with the expression, or the cognitive significance of the expression. Character is "...what is set by linguistic conventions... it is natural to think of it as meaning in the sense of what is known by the compentent language user."[7] content, on the other hand (as Kaplan uses the term), is to be equated with "what was said" via a particular utterance in a particular context; it's the sort of thing which we hold fixed when, through the use of modal and intensional operators, we want to evaluate what someone said with respect to some counterfactual situation.[8] It is this latter notion which Kaplan suggests is the one closest to the traditional notion of a proposition; it's content which is the sort of meaning which determines extension and truth value, and it's content which Kaplan thinks we normally specify in ascribing propositional attitudes to someone. Or to put it one more way: Character provides a function from contexts of utterance to contents; fix the context of utterance, and the character of the expression will determine its content. Content provides a function from circumstances of evaluation to extensions and

truth values; fix the circumstance in which a content is to be evaluated, and it determines the extension.

Now the way in which this distinction is relevant to problems with indexicals is the following: for non-indexical terms, content and character are just the same thing; that is, the character of such a term will determine the same content in each context of utterance, and its content and its character can both be identified with its "meaning" with no problems. Indexical expressions, however, are "directly referential" and have "context-sensitive" characters. To say the former is to say that the content of the expression either <u>is</u> or <u>directly specifies</u> the referent of the expression in the context of utterance; to say the latter is just to say that the content of the term varies from one context of utterance to the next -- in particular, it varies because the <u>referent</u> varies, and the referent either is or is part of the content.[9]

Let's clarify with an example; take Putnam's earlier example of my Doppelganger and I both thinking "I've got a headache". The concepts which we each associate with these expressions are the same, as Putnam would have it; what we each know by knowing the meaning of the words is the same; hence, the two thoughts or utterances have the same character. They differ in content, however. 'I' is a directly referential expression, and in <u>my</u> utterance or thought it refers to <u>me</u>, and in <u>his</u> it refers to <u>him</u>. And, at least in such a straightforward case as

this, the different ways of viewing the distinction all seem to fall into place quite nicely. As for content: what was said via my utterance, and the belief I expressed, was that I have a headache, and what he said (and believed) was that he has a headache; and in evaluating what we each said with respect to some counterfactual situation, what would matter would be whether the actual user did, in the counterfactual situation, have a headache -- whether he said anything in the counterfactual situation or not. And as for character, it would seem that cognitive significance, "what we know", and "linguistic meaning" are all invariant with respect to the two cases.

IV. The Meanings of Natural-kind terms

Keeping this means of dealing with explicit indexicals in mind, let's now return to the problem about the meanings of natural-kind terms. First, it's worth briefly examining the sort of consideration which Putnam offers in "The Meaning of Meaning" for rejecting the idea that we ought to say about these terms what we normally say about the explicit indexicals -- i.e., that they can mean the same thing but have different referents on different occasions of use. About this option, Putnam has the following to say:

> While this is the correct route to take for an absolutely indexical word like 'I', it seems incorrect for the words we have been discussing. Consider 'elm' and 'beech', for example. If these are 'switched' on Twin Earth, then surely we would

> not say that 'elm' has the same meaning on Earth and Twin Earth, even if my doppelganger's stereotype of a beech (or an 'elm', as he calls it) is identical with my stereotype of an elm. Rather, we would say that 'elm' in my doppelganger's idiolect means beech. [10]

This much here is right: one natural thing to say about such a situation is that that by 'elm', my doppelganger means beech. But notice that it's also quite natural to say such things as "By 'him', she meant George", and even "By 'that jerk', he means the guy at the end of the bar" -- even when the supplied "gloss" (i.e. 'George', 'the guy at the end of the bar') is one which our subject would not associate with the object of her thought. As Putnam acknowledges, in the case of explicit indexicals, we ought to say that the terms in question vary their referents but not their meaning on different occasions of use. In cases like those just mentioned, however, it's clear that 'means' is being used in a way which does not accord with this point. Indeed, as the second case seems to make especially clear (noting that 'the guy at the end of the bar' isn't explicitly indexical), 'means' in this context seems to be used in a way which it is interchangable with 'refers to'. Even in the case of 'I', it seems like the only thing wrong with saying "by 'I', he means himself" is that it is to say something anybody speaking the language ought to know; it's certainly true, but totally trivial.

The point of all this is just that any account of "narrow" meaning surely shouldn't be held responsible for

accounting for everthing of the form "x means (that) y" that we're inclined to hold. No account can do this, because what we're inclined to say is just plain contradictory -- compare: by 'I' I mean myself and he means himself, but the meaning of the word 'I' as I use it and as he uses it is the same. So if we're interested in clarifying the notion of meaning, we're either going to have to disregard one of these ideas, or acknowledge that 'meaning' is simply (or maybe complexly) ambiguous. In trying to develop the notion of narrow content, one is trying to sort out this ambiguity. Surely the history of science is full of cases where what looked like a single notion turned out to actually be a confusion of two (or more) distinct notions, each of which had it's own distinct theoretical interest; prominent examples are the cases of heat and temperature, and of mass and weight. So, in short, it won't do here simply to point out that we're sometimes inclined to use 'meaning' in way way which doesn't jibe with a notion of narrow meaning; rather, one would need instead to show that there isn't a natural sense of 'meaning' which is in accord with the idea that meaning is "solipsistic" in the sense desired. And so in Putnam's "beech" example, the fact that we'll say that by 'elm', the Twin Earther means beech, just isn't enough; Putnam would also need to show that there's no natural sense in which the meaning of 'elm' for me is the same as the meaning of 'elm' for my doppelganger, and this has yet to be

done.[11]

Furthermore, for cases like those at hand, it looks as though Kaplan's content/character distinction does give a natural notion of meaning which is shared by the doppelgangers if we subsume these natural-kind cases under the indexical apparatus. The content (as Kaplan uses it) of 'elm' in our idiolects is of course different, as reference is (at least one part of) the content of indexicals; however, we assign the same character to our uses of 'elm' -- where sameness of character is guaranteed at least by the use of all the same rules, concepts, perceptual stereotypes, and so on. And in this sense of meaning, surely we do mean the same thing by 'elm'.

V. Burge's Argument

So far I have suggested that if we view natural kind terms as a species of indexical terms (as Putnam suggests -- at least in this earlier article -- we do), they can be dealt with in the same way, and present no particular problem for an account of narrow meaning (yet). But Tyler Burge, particularly in his paper "Other Bodies", argues that "there is no appropriate sense in which natural kind terms like 'water' are indexical", and that hence, there is no "convenient and natural way of segregating those features of propositional attitudes that derive from the nature of a person's social and physical context, on the one hand, from those features that derive from the organism's nature, and

palpable effects of the environment on it, on the other."[12] It is to Burge's arguments that I'll now turn.

The central point in Burge's discussion of this matter is that natural kind terms should not be treated in the same way as indexicals. His reason for this is straightforward. Accepting Putnam's gloss, Burge points out that indexicals are (at least) terms which "have an extension which varies from context to context or token to token". But, he suggests, the terms under consideration don't have this property at all:

> I think it is clear that 'water', interpreted as it is in english, or as we English speakers standardly interpret it, does not shift extension from context to context in this way. (One must, of course, hold the language, or linguistic construal, fixed. Otherwise, every word will trivially count as indexical. For by the very conventionality of language, we can always imagine some context in which our word -- word form -- has a different extension.)[13]

Now we certainly don't want every word to count trivially as indexical simply because the same word form could be used in a different language with a different reference. Consider Burge's own example of the sort of "shift in extension" which we surely don't want to count as evidence for indexicality: what he says is that the analysis of natural kinds as indexical "is no more plausible than saying that 'bachelor' is indexical because it means 'whatever social role the speaker applies "bachelor" to' where 'the speaker' is allowed to shift in its application

to speakers of different linguistic communities according to context. If Indians applied 'bachelor' to all and only male hogs, it would not follow that 'bachelor' as it is used in English is indexical."[14] The question to ask here, however, is whether in order to avoid this we must "hold the language, or linguistic contrual, fixed" in the way Burge requires. For notice: the sort of shift in extension possible with natural kind terms is of a significantly different sort than that involved in Burge's 'bachelor' case -- they can vary their extension without changing the concepts, rules, and so on associated with the expression; i.e. (at least on one way way understanding the earlier notion of character) without changing their character. Indeed, such terms can shift their extensions even when all the facts about the organisms internal structure (computationally or even physically specified) are held constant -- and it's just this property which they share with explicit indexicals.

It's an interesting fact about how we individuate languages that difference in extension of (at least some) terms is reason to assume two languages. But two points should be noted here: First, the same sorts of considerations which lead one to hope for an autonomous (or "solipsistic") psychology might quite reasonably be taken to point one towards the possibility of linguistics having this same character; it's at least extremely

counterintuitive that linguistic theory should discriminate between physically indentical speakers in more or less indistinguishible environments. Secondly, and more importantly for present purposes, there is this: even if this point about languages is accepted, this is not enough to rule out a coherent notion of narrow meaning. In the same way that you and I mean the same thing by 'I' (even though by 'I', I "mean" myself and you "mean" yourself), the two doppelgangers mean the same thing by 'water' (even though the Earther "means" H2O and the Twin Earther "means" XYZ). The only difference is that one of the contextual facts in the case of 'water' one might specify context by specifying the language being spoken -- which in turn, as Burge says, may in principle fix the referents of the natural kind terms -- whereas in the case of 'I', the context must be further specified in each case. Perhaps what's been shown is that equating narrow content with linguistic meaning is somewhat misleading in the case of natural-kind terms. Rather, narrow content is a matter of the associated concepts; and there's nothing in what Burge has offered so far that shows we can't use the same sort of apparatus as we use for indexicals in "segregating those features of of propositional attitudes that derive from the nature of a person's social and physical context... from those features which derive from the organism's nature".

It's worth pointing out here how these last points

relate back to the earlier question from Putnam about 'elm' in the Twin Earth language (the on which Putnam in fact calls a "dialect of English") meaning beech, in the case where the words are switched. One way to take Putnam here is as suggesting (and indeed, something he explicitly suggests elsewhere -- cf. "Meaning Holism") that it's not just that we're inclined to say such things as "by 'elm', they mean beech", but rather that the correct translation of their word 'elm' into English is as 'beech' -- that is the way English / Twin Earth English dictionaries should be put together. Now I think that to a great degree Putnam's worries about translation are tied up with the problems of "internal" context, and will in general be put off until those problems are considered explicitly. But this much can be pointed out now. This suggestion about translation surely is just another way of making Burge's point about the individuation of languages -- fixing the language that a given natural kind term is an expression of will also fix the reference of that term. But to repeat, this point about the individuation of languages itself doesn't seem to directly impugn the notion of narrow content.

Let me turn breifly to an issue which is often not separated from the consideration of natural kind terms as a kind of indexical: that of the possible definability of indexicals in general (and so, on this line, natural kind terms) via some small class of explicit indexicals. Burge

himself considers a few options for doing this (e.g. paraphrasing 'water' as "stuff called 'water' around here") and (rightly, I think) rejects them all. The question is whether this should impugn the view that I'm pushing here. The answer is "no". Natural kind terms are like indexicals in the way I have suggested, and it is this property which seems to me to be at the heart of indexicality; but nothing I have said commits me to the view that there are only a few "primitive" indexicals and that all the seemingly indexical terms are to be accounted for as being definable by means of these "primitives" and the non-indexical expressions.

Returning to Kaplan's analysis will help in clarifying here. As he would have it, there is in a certain sense only <u>one</u> primitive indexical; that is, the <u>indexicality</u> of all indexical expressions is to be analyzed in term of what he calls the 'dthat' operator. The 'dthat' operator provides a means for constructing a rigid, directly referential, indexical expression from a non-indexical character. Thus one might characterize the meaning of 'I' as 'dthat(the current speaker / thinker)', or 'now' as 'dthat(the present time)'. Now in the sense of Kaplan's 'dthat' operator, I'm quite happy with the idea of a single "primitive" indexical being used to analyze the indexicality of all such expressions; as a piece of analytic apparatus, I think Kaplan's 'dthat' has much going for it. What I am rejecting is the idea that all indexicals are definable in terms of 'dthat' and the non-indexical vocabuary -- not for

any shortcoming in the former, but for one in the latter. Notice that in both of the above examples, the completing character for that 'dthat' operator has been provided by a definite description (i.e. 'the current speaker / thinker' and 'the present time'). Now however plausible this might be in the case of words like 'I' and 'now' (a position not without its own problems), surely there is a clear class of indexical terms for which this strategy just won't go -- those which Kaplan calls the "true demonstratives". As he puts it:

> Some of the indexicals require, in order to determine their references, an associated demonstration: typically, though not invariably, a (visual) presentation of a local object discriminated by a pointing. These indexicals are the true demonstratives, and 'that' is their paradigm.... A demonstrative without an associated demonstration is incomplete. The linguistic rules which govern the use of the true demonstratives... are not sufficient to determine their referent in all contexts of use. Something else -- an associated demonstration -- must be provided.... Among the pure indexicals are 'I', 'now', 'here' (in one sense), 'tomorrow', and others. The linguistic rules which govern their use fully determines the referent for each context. No supplementary actions or intentions are needed."[15]

So, on this account, true demonstratives (1) need to have a completing character provided, and different completing characters can be associated with a given demonstative on different occasions; and (2) the completing character is typically given via a visual presentation of a local object. Now I take it nobody would argue with (1) if

they're going accept any of the sort of story Kaplan tells. And surely (2) is intuitively quite natural and plausible. But (2) allows for at least some completing characters to be expressed non-verbally. And as this part of the paper has at least taken the surface form of a defense of certain of Putnam's older views, it's worth noting that he has (in "The Meaning of Meaning") a piece of conceptual apparatus which is introduced to play something like this very role of providing a non-verbal completing character for natural kind terms: the perceptual stereotype. Now I'm not claiming that this will do the job exactly, but it is one of the candidates which merits consideration; and it at least gives one initially plausible suggestion for what the completing character we're interested in might be, and in what sense it might be non-verbal and so undefinable via the non-indexical terms.

A last comment on the particular question of defining indexicals: One idea that has become pretty firmly entrenched in the current philosophical literature is that definitions of our expresssions in terms of other expressions in our language(s) just aren't forthcoming -- and not just in the case of special class of words (like the natural kind words), but in general. If this is right, then I take it that it shouldn't be surprising that, even given some means of "isolating" the indexical component of a natural kind term from the non-indexical part of it's character, that no short-and-easy parphrases of the "narrow"

meaning of the expression in terms of other expressions in the language are forthcoming.

Let me then tie up this part of the discussion by summarizing along with Burge. As he puts it:

> To summarize our view: The differences between Earth and Twin-Earth will affect the attributions of propositional attiudes to inhabitants of the two planets.... The differences are not to be assimilated to differences in the extensions of indexical expressions with the same constant, linguistic meaning. For the relevant terms are not indexical. The differences, rather, involve the constant context-free interpretation of the terms. Propositional attitude ascriptions which put the terms in oblique occurance will thus affect the content of the propositional attitudes. Since mental acts and states are individuated (partly) in terms of their contents, the differences between Earth and Twin-Earth include differences in the mental acts and states of their inhabitants.[16]

To similarly summarize my evalution: I'm willing to buy the first claim above -- the differences involved here affect our normal ascriptions of propositional attitudes; e.g., Adam of Earth believes that water is wet, but his doppelganger doesn't. However, I <u>do</u> think such cases should be assimilated to those of indexicals, such as the situation where Adam of Earth believes that my mother is nice, and his doppelganger doesn't. Burge's reason for rejecting this is that such differences, unlike those involving explicit indexicals, involve the "constant context-free interpretation of the terms." But Burge's "context-free" evaluation depends on fixing the

language; the suggestion being that if we don't do this, every term is "trivially indexical". But I've suggested that the mark of indexicality" we should be interested in in the present context is the potential for shifts in reference where we hold fixed the character of the expressions or thoughts -- something which ought to be guaranteed at least by fixing the (autonomously specified) cognitive (or even physical) characterization of the subject. Fixing this gives a natural alternative to fixing the languge (in this slightly pregnant sense), and avoids such cases as Burge's "bachelor / male hog" case.[17]

I think the thing to say at this point is that we've not seen anything in the discussion so far which should force us to reject the antecedently very plausible idea that in at least one sense of meaning, sameness of the organism entails sameness in meaning. The cases involving natural kind terms can be assimilated to hose involving explicit indexicals, where what we hold fixed is cognitive structure rather than "wide" meaning. But this has a serious problem in that it doesn't seem to give us a coarse enough cut. That is, even if there is a fairly clear sense in which doppelgangers mean the same thing by natural kind terms in spite of differences in reference caused by differing physical environments, it also ought to turn out -- on any natural account of meaning -- that small differences in cognitive structure are compatible with sameness of (narrow) meaning; e.g., you and I are

hardly doppelgangers, but surely whatever meaning comes to, it ought to turn out that we mean the same thing by "chair".

Exact cognitive identity is fine as a sufficient condition for sameness of narrow content, but it obviously won't do as even the roughest approximation of a neccesary one. We would like, then, to say what sort of autonomous or solipsistic sorts of considerations might play this role; and the moral of Putnam and Burge's points has been that a couple of natural candidates here -- sameness of reference and sameness of langauge (at least in the sense discussed) -- just aren't autonomous facts about the organism. But this problem of glossing over "unimportant" differences in cognitive structure is just what I earlier called the problem of the effect of <u>internal</u> context on meaning.

Before turning to an explicit examination of this problem, let me comment breifly about the sorts of remarks Burge (and Putnam, in some of his moods) makes about the effects <u>social</u> rather than <u>physical</u> context have on (wide, of course) meaning. The standard version of this is Burge's "arthritis" case: We imagine a pair of doppelgangers whose external environment differs simply in the way the word (or word form) 'arthritis' is used by certain other members of their societies. In the situation of the doppelganger here on Earth, speaking English, other people in the society -- physicians, more educated lay people -- know that arthritis is a disease of the joints

only, and hence that one cannot have arthritis in, say, the thigh. In the alien doppelganger's society, however, the physicians and educated lay people use 'arthritis' to refer to a slightly different class of afflictions; and according to their "concept of arthritis", there's nothing particularly unusual about having what they call 'arthritis' in the thigh as well as in the joints. Our present subjects, however, don't have any very well defined opinion about the possibility of having what they each refer to as 'arthritis' in areas other than the joints, but nothing in their "internalized" concept rules it out.

The case so set up, the sort of points Burge makes are much the same as those made about the effects of different physical contexts: The Earther "means" arthritis, the alien doesn't; this difference isn't to be accounted for in the way indexicals are; and so on. Now I think the general strategy for dealing with this is no different than the one taken above with the 'water' case. But this kind of case has an extra element thrown in -- it makes use of the assumption that our ubjects mean the same thing by their uses of 'arthritis' as do the physicians in their respective societies. And if "means" is taken in the wide sense (where "meaning" for natural-kind terms is "99% reference"), this seems fine. The question is whether when clarifying the notion of narrow content we either can or must make this assumption -- which once again brings us to the topic of "internal" context, and it's to this I'll now turn.

VI. Meaning and Collateral Information

Probably the most conspicuous version of this problem is posed by Putnam in his paper "Computational Psychology and Interpretation Theory"; I'll begin with a consideration of it. As for the question which Putnam sees himself as posing, "the problem is this: if the brain's semantics for its medium of representation is verificationist [or for our purposes, solipsistic] and not truth-conditional, then what happens to the notion of the "content" of a mental representation?"[18]

Something awful, we're assured; and that assurance comes primarily through the use of the following example of the two Ruritanian children:

> Imagine that there is a country somewhere on Earth called Ruritania. In the country let us imagine that there are small differences between the dialects which are spoken in the north and in the south. One of these differences is that the word "grug" means silver in the northern dialect and aluminum in the southern dialect. Imagine two children, Oscar and Elmer, who grow up in Ruritania. They are as alike in genetic construction and environment as you please, except that Oscar grows up in the south of Ruritania and Elmer grows up in the north of Ruritania. Imagine that in the north of Ruritania, for some reason, pots and pans are normally made of silver, whereas in the south of Ruritania pots and pans are normally made of aluminum. So northern children grow up knowing that pots and pans are normally made of "grug", and southern children grow up knowing that pots and pans are normally made of "grug".[19]

The first point that Putnam takes from the description

of this case is one which it would seem there's no denying: Take any account of "narrow" or "solipsistic" meaning -- any account of content such that difference in extension does not enter directly into determination of meanings; "on any such notion of content it would seem that "grug" in Oscar's mind would have the same content as "grug" in Elmer's mind. Not only would the words have the same content; any mental signs or predicate-analogues that the brain might use in its computation and that corresponded to the verbal item "grug" would have the same content at this stage."[20] The following question is then asked:

> But if the word "grug", and the mental representations that stand behind the word "grug"... have the same content at this stage, then when do they come to differ in content? By the time Oscar and Elmer have become adults, have learned foreign languages, and so on, they certainly will not have the same conception of grug.... Each of them will know many facts which serve to distinguish silver from aluminum, and "grug" in the South Ruritanian sense from "grug" in the North Ruritanian sense.[21]

So, since Oscar and Elmer have different "concepts of grug" as adults, but the same one as children, they must have changed their concept of grug along the way. But the sorts of things Oscar and Elmer learned along the way seem like just the sorts of things we would normally characterize as learning more about grug, or forming more beliefs in which their concept of grug figures. As Putnam puts it,

> ...there is no stage at which the word "grug" or the corresponding mental representation in the mind of Oscar... is ever treated as changing its reference. internally to treat a sign as changing its reference is to treat it as, in effect, a different sign. This never happens; in the internal point of view all that happens is that Oscar acquires more information about grug.... When the use of a word is modified by the continual acquisition of collateral information, without it being supposed that at any stage the word is being committed to a new extension, all that happens (in the verificationist model) is that the degree of confirmation of various sentences containing the word changes.[22]

Now the immediate conclusion that Putnam draws from this all this is a fairly mild one, and one which I'm at least initially quite inclined to accept -- it's simply that "we can have a complete description of the use of mental signs without thereby having a criterion which distinguishes changes in content of mental signs from changes in collateral information."[23] But he doesn't stop there; rather, two pages later, after quickly considering a couple of possible ways of providing such a criterion, he states what I take to be the real point of his discussion:

> Once we decide to put the reference (or rather the difference in reference) aside, and to ask whether "grug" has the same "content" in the minds of Oscar and Elmer, we have embarked on an impossible task. Far from making it easier for ourselves to decide whether representations are synonymous, we have made it impossible.... "Factoring out" differences in extension will only make a principled distinction on when there has been a change in meaning totally impossible.[23]

The problem is then this: narrow content must factor

out consideration of the sorts of differences in extension which show up in cases like the "twin earth" and "Ruritania" ones. But it's extension which is the central guide in determining when changes in meaning rather than shifts in collateral information occur. Thus, with regard to narrow content, a "principled distinction" between change of meaning and change in collateral information cannot be made. How, Putnam is asking, can we draw the line in a principled way between the meaning of "grug" for Oscar the child and Oscar the adult?

For what's to follow, let me put the point slightly differently. Let's call the totality of a representation's inferential relationships to other representations in a particular system it's conceptual role (following Field [25]). So, the conceptual role of 'water' for me now will depend on all the inferences I'm inclined to make about water, even on the basis of facts about water which are, intuitively, collateral information about water rather that facts constituitive of it's meaning. For example, I believe that there's lots of dirty water in the Charles; given that, someone who held all the other attitudes toward water that I do but failed to believe that there's lots of dirty water in the Charles would have a different conceptual role for 'water' than I do. We might then put Putnam's question like this: Which changes in conceptual role count as changes in meaning? Or, assuming that a representation's having some particular meaning is just a matter of its

having some particular kind of conceptual role What conceptual roles for a representation make it mean what it does rather than something else?

VII. Fodor's Response

What I'll do now is turn to a consideration or the sort of response made to this question by Jerry Fodor (pretty clearly a central target of Putnam's here) in some of his recent work. One place where Fodor's current view is expounded is in his paper "Narrow Content and Meaning Holism"; and here, in response the the sort of problem Putnam offers, Fodor has the following to say:

> To summarize: once you have functional role semantics you have semantic holism (and hence skepticism about the contents of propositional attitudes.)[26] ... it is notable that neither Quine, nor Putnam, nor -- to my knowledge -- anybody else, has provided serious arguments for the identification of meaning (/conceptual content) with functional role. I suspect that the main argument is simply a presumed lack of plausible alternatives. This suggests a tactic for dealing with semantic holism: namely, don't grant the theory of meaning that it presupposes.[27]

So, following this line, what we're in want of is a plausible alternative to the identification of meanings with conceptual roles; and (surprise!) Fodor has a candidate ready and waiting -- what he calls "denotational semantics". Although this view is mentioned breifly in the "Narrow Content and Meaning Holism" paper, it is in his paper

"Psychosemantics; or, Where do Truth Conditions Come From?" that this view is most fully spelled out. I would like to avoid getting into the details of the view here -- and more importantly, I think that I can while making the point I want here. So let me try to give Fodor's punch line without telling the whole joke.

On this story, what allows representations with different conceptual roles to mean the same thing is that, in spite of the possible differences in the causal chains leading to their okening, they are nonetheless both appropriately connected to the right property in the world: "...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."[28]

Now one kind of problem which Fodor admits this sort of view faces is what he calls the "thinness of slice" problem. This is, of course, just a resurfacing of the sorts of considerations that made us want a notion of <u>narrow</u> content rather than truth conditional content for the purposes of pychological explanation. As Fodor says,

> ...it's important to have a semantic theory that slices mental states thin enough; a theory which allows us to distinguish beliefs about The Morning Star from beliefs about The Evening Star, beliefs about closed triangulars from beliefs about closed trilaterals, and so forth. Now, since The Morning Star <u>is</u> the Evening Star (since all closed triangulars are closed trilaterals and vice

> versa), it is surely plausible that no purely denotational theory of content can slice mental states thin enough.[29]

Of course, something like conceptual role semantics is ideal for this task -- the critical difference in the role of 'Morning Star' and 'Evening Star' beliefs in the causation of behavior is surely a matter of the diffeent inferences one is inclined to make from such beliefs. Conceptual role semantics may have a hard time cutting slices thick enough, but thin slices are what it's made for. How do we get thin slices without recourse to conceptual role?

Easy: "The way to slice mental contents thin enough is by postulating thin properties."[30] So, suppose we want it to come out that 'closed triangle' and 'closed trilateral' have different meanings; then "one could simply take the view that the property of being a closed triangle is different from the property of being a closed trilateral." [31] But as Fodor acknowledges, this seems, at least in some cases, to be a bit much. To take an almost contemptuously familiar example: the property of being water and the property of being H20 look, at least on the face of it, to be the very same property. So (as Fodor reasonably asks himself), "how are you going to keep the thought that water is wet distinct from the thought that H20 is?" [32]

Fodor's answer is essentially the same as that given by Fred Dretske in his book knowledge and the

flow of information. Dretske's own "denotational" view on semantics is similar to Fodor's in many respects (or vice-versa, if you prefer), and particularly in the central idea that it is connection with the properties of the world rather than conceptual role which is central to meaning. As for the present problem, though, the line is this: Roughly, concepts play the role of "narrow contents" for Dretske; and so to separate coextensive concepts, he simply claims that

> the only way a system can have distinct concepts F and G, when these concepts are equivalent in one of the described ways [i.e. analytically or nomologically coextensive], is if at least one of them is complex, if one of them is built up out of conceptual elements that the other is not.... What is is impossible on the present account of things is to have two primitive concepts that are equivalent. [33]

Similarly, we have Fodor's way of putting the move:

> I think the way to fix the fatness of slice problem is to let in a moderate, restricted and well behaved amount of functional role. The point about the expressions 'water' vs. 'H20' is that, though they -- presumably -- denote the same property, the second is a complex formula built out of expressions which themseves denote hydrogen and oxygen. I do want to let into semantics -- over and above denotation -- those implications which accrue to an expression in virtue of the relations to such other expression as occur as its syntactic constituents.[34]

This may serve for cutting apart 'the morning star' from 'the evening star', and 'water' from 'H20'; trouble is, it also looks to separate 'bachelor' from 'unmarried man' as

well. Surely 'unmaried man' is, like 'H20', A "complex formula" built out of expressions which denote the distinct properties of being unmarried and being a man. And it certainly seems, at least prima facie, that it would be nice if 'bachelor' and 'unmarried man' turned out to mean the same thing on our semantic theory.

So it looks as though even the "moderate, restricted, and well behaved" bit of conceptual role allowed in here ends up slicing things up too thinly. But there's a further problem with the account at hand which is, I think, much deeper, but was glossed over in the preceeding discussion. There, we simply assumed the "obvious" candidate for the denoted property. What remains to be seen is whether, aside from the problems just noted, we can get a reasonable characterization of this denotation relation at all. So let me now turn to Fodor's characterization of this, which is to be found primarily in his paper "Psychosemantics, or, Where Do Truth-conditions Come From?" The position given here is complex and thought-provoking, and has innumerably many interesting consequences and potential problems which I'm afraid I'll just have to skirt here. What I'm interested in for present purposes is how the line in question might facilitate an avoidance of the meaning holism problems, and it's to that particular question that I'll try to confine this discussion.

VIII. Evolution and Denotation

To start with, then, Fodor tells us that "the only symbol - to - world relations that affect the semanticity of mental representation are the ones they bear to states of affairs that determine their truth values"[35] -- i.e., the only relations to the world that matter are truth conditions. But "what makes [state of affairs] S the truth condition for [mental representation] M... is that S is the entry condition for M"[36]; and "the entry condition for a mental representation M is that state of affairs such that: under conditions of normal functioning (the organism's cognitive system puts M in the yes-box iff the state of affairs obtains.)"[37]

Now the first thing to point out is how much is ridin on the "conditions of normal functioning" clause here. As Fodor readiliy points out, the entry condition for M is not just the condition(s) that is (/are) causally necessarary and sufficient for M's being put in the yes-box. Intuitively, there would seem to be two kinds of cases where the conditions responsible for M's being tokened are not M's truth/entry conditions, and which thereby need to be ruled out by the "normal functioning" proviso.

The first kind of case at least looks reasonably straightforward. Surely we would like to rule out conditions of tokening which involve such things as the intervention of neurosurgeons, hallucinogenic drugs, or shots to the head. In short, we at least want to require

something like no breakdowns of the machinery. We're pretending that there is a computational story to tell about how we go from something like stimulations to putting representations in the yes-box. Certanly any reasonable story about "normal conditions" here will require that under normal conditions, M's getting into the yes-box will be a result of the "internally correct" workings of the cognitive mechanisms; that is, it's in there because the computational system put it there, and the system isn't in any way internally malfunctioning. And it's just this sort of constraint which can keep entry conditions from including such things as the actions of brain-writing neurosurgeons.

Now it may not being entirely obvious how to characterize this "no breakdowns" state, but it looks pretty straightforward compared to the other kind of case that "normal functioning" is supposed to rule out.[38] For Fodor clearly (and with good reason) wants it to rule out cases where the interna mechanisms would not seem to be malfunctioning, but misrepresentation occurs because the external situation is not "normal" in the relevant sense. Probably the clearest (but also most farfetched) example of misrepresentation of this kind would be the standard sort of "brain-in-a-vat" case: take my brain out of my body but keep nourishing and stimulating it in the right way, and (as the fable goes) you'll be able to satisfy the "no internal malfunction" reading of normal conditions". But by doing that, you still bring about a situation where I'm

misrepresenting things. It seems to me -- and I wholehartedly believe -- that I'm sitting at my desk typing, but I'm not. We have set up conditions which are causally sufficient for my believing that I'm typing, and in which no internal malfunctions of the cognitive mechanisms have occurred; but because the conditions aren't "normal" in the relevant sense, we have what Fodor calls a "wild tokening" of the belief that I'm typing.

The question we're then faced with is this: where do we get a notion of "normal circumstance" which will do the job needed here? Or as Fodor puts it, "...if we've already used up all that [i.e. causally necessary and sufficient conditions] to establish representation, what more could be required to establish truth?"[39]

Answer: teleology.

> The distinction between normal and wild tokens rests - so far at least - on a pretty strong notion of teleology. It's only in the teleological cases that we have any way of justifying the claim that wild tokens represent the same thing that etiologically normal ones do; and it is, as we've seen, that claim on which the present story about misrepresentation rests.[40]

And how exactly is teleology supposed to support this distinction? The central idea is that "'abnormal etiology' [i.e. violation of the "normal conditions" proviso] will have to be defined with respect to the teleology of the belief-fixing (i.e. cognitive) mechanisms."[41] Defined how? Well, "a normally functioning cognitive system is one

that is doing whatever it is that cognitive systems were designed to do."[42] Or as it's put at one point:

> ...entry condtions are defined in terms of the teleology of cognitive systems (they are, for example, the conditions that such systems respect when they're oing what they were selected to do.) And the primary function of cognitive systems is, surely, to bring about coherent relations between the propositional attitudes of an organism and the states of its environment. So, for example, why does a light go on in a frog's head when a conspecific croaks? Well, because there are (cognitive) mechanisms which throw the switch just in case a certain array of acoustic energy impinges upon the frog's auditory transducers. But why are there these mechanisms? In virtue of what do they have their selection advantage? in virtue of their ability to correlate certain mental states of the frog with the presence of a croaking conspecific. So, then, what is it for the cognitive system of the frog to be functioning normally in this respect? it's for the frog's yes-box to contain a "hello, there's a croaking conspecific' token iff there's a croaking conspecific on the scene.[43]

Now in the case of the frog, it's certainly quite plausible that its cognitive mechanisms were designed (i.e. selected) for their ability (in part) to put 'there's a croaking conspecific around' in the yes-box iff there's a croaking conspecific around. Furthermore, I'm inclined to think it's quite plausible that our cognitive mechanisms (or a least some subsystem of them) were selected (in part) for their ability to, e.g., provide us with representation of the shape of the medium-sized physical object around us. Surely this is the kind of "teleology of the visual system"

that people studying the workings of the visual system talk about. The question to ask, however, is whether there's any sense in which it's plausible that our cognitive systems were selected for, say, their ability to put 'there's an airplane' in the yes-box when we're confronted with an airplane.

(Before trying to answer this, it's worth pointing out that even if the answer is yes, it may not give us what we want. We wanted from the start a notion of content which was the same for physically identical organisms. But even if this move works, what we get is a notion of content for organisms which have cognitive systems which are designed for the same purposes, and physical type identity doesn't guarantee this in principle. It's of course possible in principle to have a duplicate of me materialize from the random motions of molecules; any kind of teleological approach to content would then have such a duplicate's states have different (narrow) contents than mine.)

Of course the initial point to make here is fairly obvious: What makes (e.g.) "that's a chair" or "that's an airplane" have the truth conditions that they do can't be that chairs or airplanes were causally efficatious in the right way in the evolutionary history of the organism (natural history being what it is and all.) Our cognitve mechanisms were not selected for their ability to signal the particular properties of being a chair or being an airplane. But then how can evolution give us a grip on

"normal circumstances" for putting "that's an airplane" or "that's a chair" in the yes-box?

Fodor's response to this sort of objection:

> ...once selection has shaped a cognitive (or any other) mechanism, there are indefinitely many counterfactuals that will be true in virtue of the structure of that mechanism. Suppose that selection pressures favor organisms that con add. Then, inter alia, they favor organisms that can add 27 and 54. That can be true even though no organism ever did add 27 and 54, so that cases of doing that sum played no role in the etiology of any psychological mechanism. It is a serious misunderstanding of evolutionary theory to suppose that the explanation of a capacity by reference to selectional advantage presupposes that <u>that very capacity</u> has sometimes been exercised in the evolutionary history of the organism.[44]

So then, we should ask, what is this general capacity which (like adding in the above example) has been selected for, and of which our abilities with regard to airplanes and chair are (like adding 27 and 54) special cases which never in fact happen to arise in the selection process? As noted earlier, it's Fodor's view that "the primary function of cognitive systems is, surely, to bring about coherent relations between the propositional attitudes of an organism <u>and the states of its environment</u>." And what coherent relation is that? <u>Knowing the truth</u>.

> ...our belief / desire psychology involves us in a teleological assumption about the cognitive mechanisms; namely that they're designed to fix whatever beliefs are true. It is, I claim, only

> on this assumption that we can make sense of the semanticity of propositional attiudes.[45]

So the "normal functioning" of our cognitive mechanisms appealed to in the specification of entry (and hence truth) conditions is defined in terms of those mechanisms doing what they were selected to do; and what they were selected for was generating a perfect correspondence between states of affairs in the world and sentences in the yes-box -- i.e., for believing what's true. Of course, as Fodor acknowledges, we can in fact be fooled (e.g. by holograms) or be ignorant (e.g. because we're too far away too see). But these are, s he says, <u>accidents</u>; and "the most usual of these 'accidents' is, of course, the ailure of epistemic appropriateness onditions in virtue of acts about the causal / spatio-temporal situation of the organism. ("I couldn't see it from here,"....)"[46] Or, putting it differently: "rub our noses in the fact hat and (if we can frame the thought that) we'll com to believe that P. But, of course, for indefinitely many states of affairs... our noses are never so rubbed...."[47]

Once again I'm going to skirt some important and interesting issues that arise here, particularly (1) the potential for circularity here, and thus the failure to give a <u>naturalistic</u> account of the semantics of mental representations (a charge which Fodor tries to answer), and (2) the connections of this position with verificationism.

The issue at hand is the use of this move to avoid the meaning holism problem fr conceptual role semantics; let me then confine the points here to what I see as the failure of this particular move.

First, then, let's consider the "selection for addition" example offered; or for a first try, selection for the ability to count. It's at least plausible that the ability to count could have selection advantages (e.g. it's helpful in making sure you haven't lost one of the kids). But all of the particular counting tasks which might turn out to be efficatious in the selection process will have some upper bound -- for example, if counting is important because it allows the organism to keep track of the kids, then there won't be any selectional advantage in being able to count higher than the number of offspring had at any given time. This doesn't mean that the general ability to count couldn't have been selected for, though. It may have been that -- because of the prior structure of the organism, or even because of general facts about the biological underpinnings of cognitive mechanisms (less plausible) -- the kind of mechanism that was avialable to solve the "count to 10" task was in fact a general counting mechanism. The point is, however, that the sorts of facts in this kind of case which determine whether you get a general counter or a 10-notch tally board aren't facts about selection pressures; they're instead facts about how the species' prior structure and biological underpinnings allow

it to respond to the selection pressures.

So selection pressures favoring organisms that can count does not mean they must, inter alia, favor organisms which can count to 100. And similarly for adding: if the ability to add 27 and 54 never played a role in evolutionary history, then there are no external selection pressures which favor organisms which can add over those which can add everything except 27 and 54. Of course in this case, there some inclination to think (although it's hard to say exactly why) that the internal constraints of the machinery would favor the adder over the shmadder -- in contrast to the counting case, where if anything, the inclination seems to run the other way. But my intuitions about that aren't particularly important for present purposes. What is important is the general point: For any general capacity (like adding or counting) there will be some nite (and sometimes, as in the countng example, small) set of uses of that capacity which will, as a matter of natural history, actually be evolutionarily efficatious. And external selection pressures don't bear at all on what mechanism meets those uses or how it behaves in cases (like adding 27 and 54) which never came up. Just as data underdetermines theory, external selection pressures underdetermine cognitive mechanism.

Back to truth. Now it's ok with me if there are some representations and states of affairs for which this whole story turns out to be right. In fact, such

representations as those which describe the shape of medium-sized physical objects in our vicinity or which describe something as a human face look like reasonable candidates. There's at least some plausibility that there were direct external selection pressures which favored organisms which ere as close to omniscient as possible about the application of such descriptions. Let me call such representations (if indeed there are any such)the teleologically salient ones. The question is then of course whether there's any reason to think that -- given no direct selection pressures favoring the ability to accurately apply "chair", "airplane", or even "arthritis" -- cognitive mechanisms were selected for the ability to fix just those beliefs that are true, rather than for the ability to fix just those true beliefs representable in the vocabulary of consisting of just the teleologically salient representations. The answer to that question, as far as I can see, is no. But if that's right, then it would seem that there's no teleologically grounded notion of "normal circumstances" for the representations which are not teleologically salient, and hence no teleological story to tell about their entry/truth conditions.

But there's still an option for this line. Recall that "the entry condition for a mental representation M is that state of affairs such that: under conditions of normal

functioning (the organism's cognitive system puts M in the yes-box iff the state of affairs obtains.)" Perhaps then, although there is no teleologically grounded notion of "normal functioning" for representations like "that's a chair", we can just specify normal functioning of the cognitive machinery in general as "normal functioning (i.e. almost omniscience) for all the beliefs statable in the vocabulary of teleologically salient representations." Then, we could just say that "is a chair" has the property of being a chair as it's entry conditions just in case it's tokening is perfectly correlated with the property in circumstances which are the idealized "teleologically normal" ones -- i.e. the ones in which we have "almost omniscience" for the class of beliefs delimited above. Indeed, this may be what Fodor has in mind when he states his point by say that "representations generated in teleologically normal circumstances must be true."[48] However, as I'll try to point out now, the qualifications on this forced by what I've said so far seriously impugn the value of this line in avoiding the meaning holism problems.

Consider: First of all, recall the qualification about -- in addition to having "normally functioning" cognitive mechanisms -- having to rub our noses in the entry conditions of a representation in order to made us "almost omniscient" about it. As it's put at one point,

> In short, there are anyhow three sorts of conditions that need to be attended to in

> accounting for why a given mental symbol does (or doesn't) turn up in the yes-box: whether the entry condition of the symbol is satisfied; whether the cognitive apparatus of the organism is functioning normally in respect of the entry condition; and whether the organism is appropriately situated is respect of the state of affairs that satisfies the entry condition.... The omniscience claim is in force only when all three sorts of constraints are simutaneously satisfied.[49]

Now in the case of the teleologically salient representations, it seems at least plausible that rubbing our noses in the entry conditions might simply amount to something like our giving it a good inspection in favorable perceptual conditions (good light, etc.). One plausible way to view it (which Fodor would however reject) might be in terms of getting all the possible epistemic access to the entry conditions that could have been had in the evolutionary environment -- roughly: no microscopes, but you can look and touch all you like. The question to ask, however, is what such a nose-rubbing might amount to in the case of such non-teleologically salient properties as arthritis. In particular, would such a nose-rubbing include rubbing one's nose in social and linguistic facts or not?

If so, then it looks as though the game has been given up to Putnam and Burge. If (a) the content of the representation 'arthritis' depends on what it denotes, and (b) what it denotes is a matter of what I'd take it to denote given all the relevant evidence (including such things as finding out how the experts use the term and,

say, discovering that Putnam's thesis of the linguistic division of labor is true), then the content of my mental representation 'arthritis' is only determined against that background of social and linguistic facts. A doppelganger of mine in a society where the experts used the term differently, or where other social or linguistic facts affecting the reference of the term were different, would not have the same content for his representation. Or to return to Putnan's "grug" case: Oscar and Elmer would as children, in spite of their identical makeup, already have different contents for 'grug'. But the point of the notion of narrow content was <u>just</u> to rule this sort of thing out. If narrow content is a matter of denotation under ideal conditions, it had better turn out that 'grug' denotes the same thing under that idealization whether it's used by Oscar or Elmer. And furthermore, it had better turn out that 'water' denotes the same thing for a pair of dopplegangers Putnam's original "H20/<u>XYZ</u>" case, in spite of the fact that a different substance played the appropriate role in the two subjects' evolutionary histories.

But if such social and linguistic facts are not among those which our noses must be rubbed in to satisfy the conditions for "almost omniscience", then it looks as if we're going to have the same kind of "slicing too thin" problem that we had with conceptual role semantics. For without either social and linguistic constraints <u>or</u> a

teleological grounding for the non-teleologically salient representations, there will be what look like lots of different entry conditions for (what was at least intuitively) a given representation, and no way to pick just those which are the truth conditions -- just as there are lots of different conceptual roles for it, and no way to pick just those with the same meaning.

This is just the sort of point which an example like Burge's "arthritis" case brings out. Consider the case of two people, both of whom have the same "perceptual stereotype" for arthritic pain, both of whom believe some of the folk wisdom connected with arthritis (e.g., it's more common among old people, it's sometimes worse in the cold, asprin helps, etc.); but one (rightly) thinks you can only get it in the joints, and the other doesn't. Now if in "rubbing their noses" in the world, we <u>don't</u> rub their noses in things like the fact that experts in their society use the term in a certain way, it looks like it's going to turn out that the idealized correlations of their representations 'arthritis' will be with different properties. Even if it's plausible to think that the idealized correlation for the guy who thinks you can only get arthritis in the joints is really with the property of being arthritis (which is pretty questionable), it's surely not plausible to think that this will be the correlation for the guy who <u>doesn't</u> think that. For him, <u>maybe</u> the property picked out will be some disjunctive one (i.e.

"arthritis or ..."), but without social or linguistic facts, or the sorts of facts about microphysical structure which would defeat the purpose of a notion of narrow content, it looks as though the denotations will be at least slightly different. But this is then just the counterpart of the problem we had with conceptual role semantics: When are different (idealized denotations / conceptual roles) similar enough for sameness of narrow content?

Let me make a last negative point about the proposal under consideration before closing with some remarks on where all this leaves us. One of the things that any useful idealization needs to do is to resemble the real case adequately in order to give useful explanations of the real cases and (closely related) to justify the idea that this idealization really is an idealization of the real cases that we're worried about. But of course, they don't always succeed: idealization in economics to perfectly rational market agents, and in political science to perfectly well-informed voters, are examples where the distance between real and ideal is great enough to strip the idealization of much explanatory value. And although I think that idealization to some sort of "almost omniscience" for some mental representations (perhaps even the teleologically salient ones) may in fact be a methodologically fruitful one[50], I see no reason to believe this for representations which are, intuitively speaking, as "non-observational" as "grug" or "arthritis".

IX. Concluding Remarks: Conceptual Role Revisited

Where have we then been left? Well, in spite of the failure of the denotational approach to give an account of psychological content in general, I think there's a good bit to be taken from the failure (and perhaps, partial success) of the view. There are, I think, two main positive points to be taken from all this: one is the possibility of a denotational account of some of our mental representations; the other is a suggestion about the importance of idealization in this problem. I'll take these in order.

As for giving an account of the semantics of mental representations which is partly denotational in the way suggested, I think the possibilities are quite open. I didn't give any argument against the possibility of a denotational account of the teleologically salient representations, and I can't think of any roughly non-empirical argument against this idea. The real roadblock for such a story is making persuasive the idea that at least some subset of our mental representations are in fact teleologically salient in a robust enough sense -- i.e., that the mechanisms for tokening such representations are "hard-wired" by the evolutionary process. Fodor has elsewhere[51] made a persuasive case that some parts of our cognitive apparatus are not as plastic as is often suggested. If this is right (and it is

at least an open empirical claim), and some cognitive "modules" (like the visual input processor) are explicitly "hard-wired" by evolution to perform certain representational tasks, then the representations of such modules would be leading candidates for the kind of denotational approach which has been discussed here.[52]

Idealization. the other important point which might be taken from the foregoing discussion is that the notion of idealization seems like the best idea available for getting around the meaning holism problem. And although idealization to teleologically defined "normal circumstances" doesn't do all the work Fodor want from it, it does suggest another kind of approach which might be taken. Recall that the question which Putanm has faced us with is this: How much change in conceptual role can you have before you get a change in content of representation rather than just a change in belief? And the problem was, there seemed to be no way that that question could be answered.

So, said the zen master, unask the question. Admittedly, there isn't any way to draw the line between change of meaning and change of collateral information. So don't do it. Idealization gives us a way to not be so bothered by this purported failure. The fact to focus on is that it's generally true that there's no line to be drawn between cases which can be subsumed under some scientific idealization and those which can't. There's

no drawing the line between say, gases which are sufficiently close to an ideal (or "perfect" if you prefer) gas for ideal gas laws to explain their behavior and those which aren't sufficiently close. And similarly: there may be no saying how close the conceptual roles of two representations have to be to subsume them under the same intentional explanation; but that doesn't mean intentional explantion is to be left for the poets.[53]

In short, the idea here is to idealize to some particular set of conceptual roles, and stop worrying that there's no line to be drawn between those conceptual roles which have the same meaning as the ideal ones and those which don't. Now before considering quickly a couple of drawbacks to this outlook, let me suggest one strength in the present context. If there are in fact, as suggested above, teleologically interpreted representations which lie at the interface between perceptual (and perhaps motor-control) modules and the rest of the cognitive system, then these might provide some class of "semantic primitives" which would have (at least some of) their semantic properties in virtue of something other than just their inferential relations to other representations.[54]

Now one kind of problem with this approach is that there doesn't seem to be any obvious preferred idealization. In the case of idealization in other sciences, there generally does look to be such a preferred case -- intuitively, one where some of the variables drop

out. You get to, e.g., stop worrying about exactly how elastic the particle is, how friction-free the plane is, or how often the consumer really reads the label. With meanings, though, it's hard to see how we could view one conceptual role as the one which really had a certain meaning, whereas the others didn't.[55] Omniscience might have given such a preferred case, but its problems have already been pointed out. In fact, in our actual, everyday application of intentional ascription and explanation, there's at least some plausibility that it's our own case that we use as the idealization.[56]

But why should we be so bothered by this? We have, in terms of Field's conceptual role, a perfectly well-defined notion of exact sameness and difference of meaning. There's no obvous reason why which full conceptual schema we actually choose as the ideal one should be constrained by anything other than pragmatic success of the explanory system. Indeed, such a choice may be reasonably viewed as simply analogous to the choice of a coordinate system. If we can't find any choice which is useful, then we should start to worry. But look: just use mine -- it's okay with me. If we really do use our own cases in practice, at least you'll have as an idealization a characterization of meanings which facilitates (I blush to say) a fair bit of pragmatic success in explanation and prediction.

Might not be ideal, but it'll have to do. Know what I mean?

NOTES

[1] See Stitch (1983), especially chapter 8. I think that something like this has been assumed by a lot of people (perhaps implicitly); another place where this sort of line is pushed explicitly is in Fodor (1980a).

[2] This is perhaps the central point of my arguments against Searle part one of this thesis, "The Chemistry of Intrinsic Intentionality."

[3] In Putnam (1975).

[4] Putnam (1975) pp. 233-4.

[5] Putnam (1975) p. 234.

[6] I take it that this is an expansion of Kaplan's line which he himself would in fact resist. Although he doesn't explicitly discuss natural kind terms, his position on proper names (in Kaplan (unpublished), ch. 22) suggests what sort of line he might be inclined to take on this subject -- and it's not the one I'm suggesting.

[6] Kaplan, p. 25.

[7] See Kaplan, pp. 19-24 for the initial exposition of his notion of content.

[8] For present purposes, I'll just say, with Kaplan, that the content of an indexical is its referent, although I

think there's good reason to say instead (as Putnam seems to) that the referent is just one component of the content of an indexical term -- another component being something like the character.

[10] Putnam (1975) pp. 245-6.

[11] Another way to take what Putnam has to say here is to take him as making a point about how we would translate between English and Twin Earth English. (Whether he's suggesting this point here or not, it's certainly one he makes elsewhere -- e.g., in his paper "Meaning Holism".) I'll get to this later.

[12] Burge (1982), p. 103.

[13] Burge, p. 103.

[14] Burge, p. 105.

[15] Kaplan, pp. 9-10.

[16] Burge, p. 107.

[17] There's still the difference between natural-kind terms and the standard indexicals with respect to which features of external contexts they are sensitive to in fixing referents. There's at least some plausibility that in the case of the standard indexicals, all that matters is what's around at the moment to, roughly, be pointed at. But with the natural kind terms, my history (among other things)

matters. Just because I may be around XYZ now doesn't make 'water' in my idiolect refer to it. Now although this surely makes it hard to say what the referents are in any particular case, I'm not sure why this should make any qualitative difference.

[18] Putnam, (1984) p. 7.

[19] Putnam (1984) p. 7.

[20] Putnam (1984), p. 9.

[21] Putnam (1984), p. 9.

[22] Putnam (1984), pp. 9-10.

[23] Putnam (1984), p. 10.

[24] Putnam (1984), p. 13.

[25] See Field (1977).

[26] Fodor, (unpublished-a), pp. 28-9.

[27] Fodor (unpublished-a), p. 27.

[28] Fodor (unpublished-a), p. 28.

[29] Fodor (unpublished-a), p. 30.

[30] Fodor (unpublished-a), p. 28.

[31] Fodor (unpublished-a), p. 28.

[32] Fodor (unpublished-a), p. 32.

[33] Dretske (1981), pp. 215-6.

[34] Fodor (unpublished-a), p. 33.

[35] Fodor, (unpublished-b), p. 14.

[36] Fodor (unpublished-b), p. 44.

[37] Fodor (unpublished-b), p. 37.

[38] Fodor makes this distinction himself on p. 60 of the "Psychosemantics" article, where he is talking about the cognitive system of a frog: "In one sense, a cognitive system is functioning normally whenever it does whatever intact cognitive systems do. In that sense, fly-detector neurons are functioning normally when they respond to moving spots. In another sense, however, a normally functioning cognitive system is one that is doing whatever it is that cognitve systems were <u>designed</u> to do. It's true that, in <u>this</u> sense, false positives to spots are abnormal."

[39] Fodor (forthcoming), p. 17.

[40] Fodor (forthcoming), p. 22.

[41] Fodor (forthcoming), p. 23.

[42] Fodor (unpublished-b), p. 60.

[43] Fodor (unpublished-b), pp. 38-9.

[44] Fodor (unpublished-b), pp. 50-1.

[45] Fodor (unpublished-b), pp. 58.

[46] Fodor (unpublished-b), pp. 51.

[47] Fodor (unpublished-b), pp. 50.

[48] Fodor (forthcoming), p. 22.

[49] Fodor (unpublished-b), pp. 53.

[50] It's in fact plausible to think that what limited successes behaviorism did have depended on something very much like this sort of idealization.

[51] See Fodor (1983).

[52] There may also be the possibility that such hard-wiring is done not through the evolutionary process, but through the <u>developmental</u> one. I in fact think kind of approach may hold some promise, in spite of such problems as that of distinguishing the sort of developmental process from cognitive processes in general. A topic for another time.

[53] This is, of course, a use of standard move #1 in philosophy of psychology: "Sure, it's a problem -- but it's a problem for 'legitimate' sciences too."

[54] The importance of this sort of "grounding" of conceptual roles is also discussed in the other two sections of this thesis.

[55] As Ned Block has pointed out to me, it may be even

worse than this. In the other good idealization cases, we have something like a similarity metric for saying what counts as being closer to the idealization -- e.g., in the case of gases, smaller particles, etc. However, in the case of conceptual roles, it's hard to see how to characterize such a thing.

[56] This point is made nicely in Stitch (1982).

PART 3:

HUSSERLIAN BRACKETING

IN

COGNITIVE SCIENCE

I. Introduction

In this paper, I will try to examine some connections between two kinds of approaches to thinking about the intentionality -- the aboutness, directedness, or semantic relatedness -- of consciousness: on the one hand, that of roughly Husserlian phenomenology; and on the other, the sort of approach which is suggested by the currently somewhat fashionable view of the mind as a computational system. In doing this, I will focus on the comments and arguments in the area which have been put forth by Hubert Dreyfus, and will emphasize in particular the statement of his position on this issue which is given in his in introduction to his anthology, Husserl, Intentionality, and Cognitive Science. I should say here at the start that this isn't really a paper on Husserl. Rather than trying to engage in any kind of Husserl scholarship, I will for the most part be concerned with Husserlian phenomenology as viewed by Dreyfus. I will, however, on occasion try to point out ways in which a slightly different understanding of what Husserl himself has

to say may help in avoiding some of the problems suggested by Dreyfus.

The first thing that I want to try to put aside is a certain kind of worry about consciousness taken in the slightly mysterious and mystical sense it often is. What I would like to do for present puposes is not to worry about the "riddle of conciousness", whatever that is. If there's a riddle of consciousness, not only do I not know the punch line, I don't even know the straight man's part. In not concerning myself with this, I'm simply going to follow Dreyfus' lead. In his attempt to avoid these thorny problems, and in trying to bring out the interesting connection he sees between Husserlian phenomenology and cognitivism, Dreyfus makes the following claim:

> ...for Husserl, like Kant, the notion of mental activity is so broadened that it does not require consciousness at all. Indeed, Kant and Husserl are precursors of cognitivism precisely because their rules operate like programs totally independently of the awareness of a conscious subject.[1]

Whether this is in fact true or not is, I think, somewhat up for grabs. However, for the most part, I'll simply buy the line that Dreyfus is giving here, and focus instead on his independent worries about the relationship between the two views at hand.

II. Bracketing and Methodological Solipsism

Rather than then telling us something about the somewhat mystical nature of consciousness, Dreyfus sees "what he [Husserl] considered his most important discovery" as that of "the special realm of entities revealed by the transcendental phenomenological reduction."[2] Now the transcendental phenomenological reduction is a "reduction" of the subject matter of a discipline -- phenomenology, or phenomenological psychology -- to just that which is available to reflection once all knowledge of the real, external world has been "put aside" or -- as Husserl puts it -- "bracketed." This act of "bracketing" -- which Husserl calls the "epoche" (abstention) -- is not a denial of the existence of the real world, or a reduction or redefinition of claims about the world in terms of what's left after the epoche, but is simply a bit of the methodology of phenomenology. In bracketing, we are told, "I do not then deny this "world", as though I were a sophist, I do not doubt that it is there as though I were a sceptic; but I use the "phenomenological" epoche, which completely bars me from using any judgement which concerns spatio-temporal existence (dasien)...."[3] Or as David Woodruff Smith and Ronald McIntyre put it in their book, Husserl and Intentionality: A Study in Mind, Meaning and Language:

> A "reduction" in Husserl's sense is a methodological device for "reducing", or narrowing down, the scope of one's inquiry. Importantly, then, Husserl's reductions are not ontological reductions, whereby entities of one category are defined or eliminated in terms of entities of some other category (as some have sought to reduce physical objects to sense-data, or minds to bodies, or values to facts, and so on). Rather, the purpose of Husserl's reductions is that of succesively delimiting the subject matter of phenomenology.[4]

Now as for the actual practice or use of bracketing itself, I won't have much to say. What I do want to focus on here is the nature of what Dreyfus calls "the special realm of entities revealed by the transcendental phenomenological reduction." This realm of entities and the operations defined over them form, for Husserl, the subject matter of phenomenological psychology; and what is critical about these entities is, Dreyfus tells us, the following:

> What is essential to phenomenological psychology is that there be an autonomous realm whose rule-like operation can be understood without reference to the activity of the brain, without asking whether anything is actually causally affecting our sense organs, without deciding whether the natural world is or is not the way science tells us it is, without asking whether any of our intentional states are actually satisfied, and, most generally, without taking a stand on whether anything at all exists for our mental states to be satisfied by.[5]

It is this central use of the notion of an autonomous, rule governed realm of mental operations which is the critical tie between Husserl's phenomenology and the cognitivistic outlook. The independence of the taxanomy of

mental states from considerations involving the external world -- the idea that mental states are what they are independent of what the external world is like -- is embodied for Husserl in the notion of bracketing. But this very same idea, Dreyfus says, is central to the cognitivistic outlook, and is captured there by the notion of methodological solipsism.

The idea of methodological solipsism is most clearly spelled out in Jerry Fodor's article "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", and it is Fodor's version of the view which Dreyfus considers explicitly. Thus, although I'm not in absolute agreement with Fodor's characterization of this outlook, it's nonetheless the obvious place to start considering the view. For Fodor, methodological solipsism is a requirement placed on psychology by another closely related supposition -- the formality condition. The formality condition, when added to the thesis that mental states and processes are representational -- i.e. that "all such states can be viewed as relations to representations and all such processes as operations defined on representations"[6] -- gives what Fodor calls the computational theory of mind.

What then is the constraint which the formality condition places on the representational theory of the mind? As Fodor puts it:

> Formal operations are the ones that are specified without reference to such semantic properties of representations as, for example, truth, reference, and meaning.... formal operations apply in terms of the, as it were, shapes of the objects in their domains....[7] ...the formality condition, viewed in this context, is tantamount to a sort of methodological solipsism. If mental processes are formal, then they have access only to the formal properties of such representations of the environment as the senses provide. Hence, they have no access to the semantic properties of such representations, including the property of being true, of having referents, or, indeed, the property of being representations of the environment.[8]

However, not all of semantics is left behind by the formality condition; for, we are told,

> ...the content of a representation is a (type) individuating feature of mental states.... But, now, if the computational theory of mind is true (and if, as we may assume, content is a semantical notion par excellance) it follows that content alone cannot distinguish thoughts. More exactly, the computational theory of the mind requires that two thoughts can be distinct in content only if they can be identified with relations to formally distinct representations.[9]

So methodological solipsism (or, if you prefer, the formality condition), like Husserl's bracketing, makes the assumption that that the "external" properties of our intentional states, such as what particular real object they happen to be about, or whether or not they happen to be true, are outside the scope of what psychology should look at. And similarly, Dreyfus says, "this bracketing of the concerns of naturalism, along with the implicit denial of the causal component of reference, makes Husserl a

methodological solipsist."[10] Indeed, Dreyfus takes Husserl's move from his earlier (pre-transcendental reduction) views in Logical Investigations to his post-reduction views in Ideas and later works to essentially the move of adding the formality condition to his representational theory of mind. As he puts it:

> Husserl's theory of intentionality developed through two stages. The first stage corresponds exactly to what Jerry Fodor, in his article on methodological solipsism, calls the representational theory of mind; and, we shall argue, the second stage may be linked to what Fodor calls the computational theory of representations.[11]

Now I take it there is at least some initial inclination to think that the same things are supposed to be "bracketed" by, one the one hand, the formality condition, and on the other, the transcendental reduction. For surely both require that claims about the existence of particular external objects, the success or failure of attempts to refer, and the truth or falsity of representations must be bracketed; but that what makes a particular representation the intentional type that it is is the sort of thing which will not be bracketed. Furthermore, the two views would seem to share at least two central motivations for making the methological reduction of subject matter via bracketing and the formality condition.

One is the obvious one -- a (tentative, anyway) acceptance of the roughly Cartesian intuition that our

mental states could have been exactly as they are regardless of the state of or even existence of the external world. It's hard to see exactly how to argue for this, but it's certainly something that people typically take to be not only plausible but obvious. The second shared motivation is a little less obvious, but perhaps more important. This is the desire to get a science of the mind which is in a certain sense "presuppositionless". Now for Husserl, the sense in which the science of the mind is supposed to be "presuppositionless" is often taken to be that of something like standard epistemological foundationalism -- i.e. depending on only the "indubitably given foundations" of, presumably, something like sense-data. I'm inclined to reject this way of viewing Husserl's epistemology (particularly as it appears in his later works), but I won't argue the point here. Suffice it instead to point out that Husserl seems to be particularly interested in not presupposing any other science or body of scientific knowledge. The references to the bracketing of scientific knowledge in particular appear constantly in Husserl's writings; e.g., "Thus sciences which relate me to this natural world... though I am far from any thought of objecting to them in the least degree, I disconnect them all... no one of them serves me for a foundation".[12]

It is this latter way of being "presuppositionless" --

i.e. not presupposing some other sciences -- which is a central motivation for methological solipsism in cognitive science as well. For, the moral of the recent literature on meaning and the fixation of reference goes, the meanings and extensions of at least some terms depend on facts about "hidden essences" of the things we refer to, and what science can tell us about them. According to the post 1970's conventional wisdom, 'water' refers to H20 and 'salt' refers to NaCl, whether the user knows any chemistry or not. Hence, whether a thought "water is wet" is about water or not "depends on whether it's about H20; and whether it's about H20 depends on 'how science turns out' -- viz., on what chemistry is true."[13] So if individuation of contents (and hence mental states) is done via their "external" semantic features (like their referents), we won't be able to type-individuate mental states without finishing up our chemistry (and presumably the rest of our sciences) first. As Fodor puts it: "No doubt it's all right to have a research strategy that says 'wait awhile'. But who wants to wait forever?"[14] Thus, since not honoring the formality condition seems to make the project of intentional psychology hopeless, all we can do is hope for a psychology which does honor it -- one which, as Husserl puts it, "puts out of action" these naturalistic notions.

What I'd like to do now is tentatively accept Dreyfus' suggestion that these central notions in the two views

really do come to the more or less the same thing: that both views are adopting the same central construct of an "autonomous" realm of rule-governed processes which are taxonomized idependently of their relationships (causal or semantic) to the external world -- an external world which they are nonetheless in fact semantically directed upon. The question I'll now turn to: What's wrong with that?

III. What's wrong with bracketing, part 1: Meaning Holism

The central problem for the notion of bracketing or methodological solipsism is what's sometimes called the problem of the background. The idea is this: A representation doesn't have the content that it does singly or in any way which is independent of the other representations in the same network. There is, to use Husserl's term, an "infinite horizon" of meanings and intentional states against which each representational content -- or "noema" -- functions; and without that background, the representational state does not have the same content.

One way to view the problem is this: Once bracketing or the formality condition is adopted, what's left are the relationships <u>between</u> the representational states. Roughly, the sort of content which is left looks as though it must be determined by formally characterizable (loosely, <u>syntactic</u>) interactions relations to the other repesentational states (including perceptual ones). In the

current literature in the philosophy of mind, the idea that of conceptual role semantics. What gives a representation its conceptual role is its connections within the conceptual or inferential network. Similarly, what makes a noema the one that it is is the fact that it connects (or "synthesizes") representations: For example, what makes a noema that of "house" is that it synthesizes our beliefs (e.g. the belief that houses are often wooden), our perceptual presentations (e.g. the appearance of the front of a house), and our expectations (e.g. that a house won't usually disappear instantaneously).

Problem: Which of the connections in the representational network are constituitive of a representation's content; i.e. which synthesizing connections are essential to being that noema? Of course, my representation "house" is directly inferentially tied to things as idiosyncratic as memories of feeling guilty about breaking a window on the green one inhabited by Mrs. Elhart which was next door to my parents' house; and indirectly tied to my beliefs about anything you like -- say, moral philosophy. The problem is that there doesn't look to be any way in principle to separate these connections from connections which might seen to be more essential to the content of the representation.

Now I'm inclined to think that this is a deep problem, and not one easily solved. If beliefs about anything you like can affect beliefs about anything else -- i.e. if

epistemological holism is true -- and if all we have for psychological content are the (roughly) epistemological or inferential relationships between representations, then it's hard to see how we're going to avoid meaning holism.[15] But for present purposes (and maybe in general), its not obvious that this is such a problem. If we simply admit that the horizon of each noema is infinite, and that the representational content or conceptual role does depend on the totality of a representations inferential role within a system, what do we lose? We still have a notion of meaning which doesn't have to "presuppose" (in the above-mentioned more specific sense) any science. What we don't have is, first, a notion of the content of a representaion which is coarse enough to include different people, or even the same person over changes in beliefs. This is, I think, a problem for Husserl's project of "eidetic reduction" -- the reduction to essences. And second, the task of spelling out any given noema or conceptual role will be, to say the least, monumental. As Dreyfus says,

> During twenty-five years of trying to spell out the components of the noema of everyday objects, Husserl found that he had to include more and more of a subject's common-sense understanding of the everyday world.... he concluded... that phenomenology was an "infinite task".[16]

However, one might hope that it's one we might actually be able to start. We all know that the inferential

connections of, say, 'water' to 'liquid' and 'drinkable' are, in some sense, more important and intimate than it's connection to 'baseball' and 'quark'. That's at least a place to start. In any case, an "infinite task" may not be so hot, but it surely seems better than one you have to wait forever to start.

I'm not trying to suggest this isn't a deep and troubling problem. However, I do want to suggest that (1) for the reasons above, it might not be as bad as it first looks; (2) there may be some kind of way around it (see the end of part 2 of this thesis for some suggestions about this); (3) it's not any new or special problem that comes up from the parallel between cognitivism and Husserl, but rather one which Quine (and Putnam in some of his moods) has been throwing around for thirty years; and most importantly for present purposes, (4) it doesn't seem to be the one which Dreyfus is actually trying to get at himself. It's the problem which Dreyfus is instead trying to present that I'll now turn to.

IV. What's wrong with bracketing, part 2: skills

The problem which Dreyfus is instead worried about is that representational structures alone are inadequate for a theory of mind -- or, as he likes to say, for an account of human understanding. For, as he says in his criticism of "Methodological Solipsism", the cognitivist (Fodor in this case), not to mention Husserl, "needs an account of how one

determines the quality of an act, that is, a theory of taking-to-be-true, even within his solipsistic method.[17] Roughly, the claim is just that both theories of mind need to give some kind of account of fixation of belief -- and in particular, beliefs which themselves involve semantic notions like truth and reference. But the problem is not just that such the fixation of such beliefs gives the best kind of example of the effects of epistemological holism; rather, it's rather (Dreyfus claims) that the sorts of epistemological background agaist which justification is made is not simply representational. The epistemological holism of belief fixation here shows that if (as it should) a theory of mind is to give an account of such belief fixation, it's going to have to be "committed to capturing the intentional structure and mental operations involved in all forms of intelligent behavior, even pragmatic, contextual interactions with objects and people in the physical and social world.[18] But then:

> The crucial question becomes: Can the taken-for-granted everyday common-sense background presupposed in assigning satisfaction conditions to every intentional state be treated as a belief system which can be analyzed in terms of the intentional content of each of its constituent beliefs? Or is the background rather a combination of skills, practices, discriminations, etc., which are not intentional states, and so, a fortiori, do not have the some of intentional content which could be explicated in terms of formal rules?[19]

Husserl, we're told, must (and does) take the former

option; the realm of abstract rules is all he has left after bracketing, and so if he is to give an account of belief fixation, of taking-to-be-true and taking-to-refer, it must be given in just these terms. Dreyfus' statement of the point here:

> Husserl thus accepts Heidegger's argument that each noema functions only against the practical horizon of the life-world, and then claims that these background practices themselves are really a set of "sedimented" background assumptions, each of which has its own noematic content, which need only be "reactivated" by the phenomenologist.[20]

The question to ask here, of course, is why Dreyfus thinks (he says, a la Heidegger) the background against which mental activity takes place must be taken to have a non-representational character? And the answer, as already implied above, is this: The contents of our intentional states depend intrinsically on the relationships they bear to skills (including, importantly, perceptual ones -- i.e. abilities to discriminate); and these skills cannot themselves be considered representational states or processes.

Now on one way of understanding Dreyfus' talk about skills here, it's just plain hard to see why we should believe this last point. This way is to emphasize the Heideggerian talk about the "socially organized nexus" which he tends to slide into in this context; e.g.:

> When we use a piece of equipment like a hammer, Heidegger claims, we actualize a bodily skill (which cannot be represented in the mind) in the context of a socially organized nexus of equipment, purposes, and human roles (which cannot be represented as a set of facts). This context and our everyday ways of skillful coping in it are not something we know but, as part of our socialization, form the way that we are.[21]

He is, it would seem, glossing together a couple of disparate points here. One is the claim that the "context of a socially organized nexus... cannot be represented as a set of facts". Now the earlier points about epistemological and meaning holism may show us that this social context is (almost) infinitely complex, but nothing so far has shown us that it's not representable as an (admittedly unbeleivably complex) set or network of "facts" or contentful representations. A big set is still a set. What's needed for Dreyfus' point is to show that skills, as they enter into the total intentional network (the "lebenswelt") are essentially bound up with something other than representational states and processes. So what we need to look for is something problematic about skills other than the fact that they, like everything else, are involved in the holistic network. So let's look.

There would seem to be two distinct ways of understanding Dreyfus' worries here skills: One is as a concern about the feasability of giving an explanation of how we accomplish skillful behavior (including managing perceptual tasks like that of identifying an object as being, say, a chair) in terms of computations and

representations; the other is as a worry about what might be called the essentially "world directed" nature of perception. What I'll do is finish this section by talking breifly about the first of these, and then focus on the second in the next section.

The first problem -- the suggestion that it's somehow unreasonable to try to give an account of how we accomplish produce skilful behavior from within a computational framework -- seems clearly to be the sort of question to be answered by theory-building in cognitive science rather than by philosophical reflections. It is, I take it, a substantive issue whether skills or perception (or for that matter, any kind of mental process) can reasonably be explained at a at a higher level of abstraction -- that of computations and representations -- than the that of some non-computational science -- say, neurophysiology. Perhaps it can't, and no computational theory of such processes is forthcoming (or true).[22] But this will depend on How Research Turns Out. The present question is whether, pre-research, we have some good reason to think that a computational account of (say) perception should be ruled out.

As far as I can see, Dreyfus really doesn't give much in the way of reasons for believing this. What he does have to offer here, though, seems to fall into two categories. One, the less interesting, consists of reflections on how armchair considerations in <u>favor</u> of thinking of skills

as rule-governed really aren't so powerful. For example (in talking about a case which, I'm afraid, doesn't apply too well to perception):

> ...when a skilled performer is asked how he does what he does, he often tell you the only thing he knows, viz., the sequence of operations he once followed in acquiring the skill; but that does not mean he now follows those steps or any others, and the flexibility and success of the skill suggests that rules no longer play a functional role.[23]

Here, the suggestion is that consciously accessible rules or procedures may well play a role in the learning of a skill, but shouldn't be taken to continue to play a role once that skill has been thoughly learned. Learning a dance is a good example of this: after learning a dance, we at least cease to be conscious of the procedures (e.g. "first move the right foot back, then the left foot in...) we used in learning.[24] As far as this sort of introspective "evidence" fo the rule-governed natur of skills goes, however, I'm perfectly inclined to agree with Dreyfus on it's status. How we <u>say</u> we do things can often be confabulation; what's critical for deciding about rules and representations as the basis of skillful behavior isn't intospective evidence, but (as I said earlier) the success of scientific theory-building from with this perpsective.

The second (closely related) category of considerations offered by Dreyfus against the pretheoretical plausibility of a rule-governed account of skills revolve around the differences between skill-governing processes and

conscious inferential processes. If we use rules and representations in, say, the solving of perceptual tasks, they at least aren't ones to which we have conscious access; we don't have introspective access to how we solve these problems, even of the restricted sort we have to how we consciously accomplish tasks or solve problems by "figuring them out." But why should this bother us? Why shouldn't we think that some of the rules and representations used in mental processing are not consciously accessible? This is not only perfectly reasonable for (and commonly used by) cognitivists, but -- importantly for present purposes -- also for Husserl, given Dreyfus' characterization of his account being centrally concerned with the supposition of a realm of formal rules involved in mental processes rather than with the "mystical realm" of consciousness (see section (I) of this paper).[25]

V. What's wrong with bracketing, part 3: world-directedness of perception

Aside from these kinds of concerns then, how does bracketing affect how we view the role of perception in mental activity? To start with, as Dreyfus rightly points out in his article "Husserl's Perceptual Noema",

> ...Husserl must, therefore [given bracketing], abandon an account of outer intuition. He must treat perception as referentially opaque and confine himself to what we take there to be rather than what is given. He can study the conditions

> of the possibility of evidence, confirmation, etc., but never it's actuality.[26]

Surely this is right. Once we have bracketed the world -- or taken the solipsistic turn -- perceptual states and representations must, like any mental states and representations, be seen as playing their role in mental processes purely in virtue of their formal or syntactic properties. Just as truth and falsity of conceptualized beliefs are abstracted from, so is veridicality of perception. As far as a theory of mental activity goes, perfect hallucination is as good as the real thing. It should, however, be once again emphasized that this does not mean that Husserl doesn't think there's a difference between perfect hallucination and veridical perception. It's just that this difference isn't a difference for phenomenology (or phenomenological psychology) to be concerned with.

As we saw earlier, what was left of the semantic properties of mental states after reduction was, on Fodor's statement, the content of the representations individuated opaquely, and on Husserl's, the act's intentionality or "directedness". As Dreyfus says, in ideas Husserl "argued that an act of consciousness... has intentionality only by virtue of an 'abstract form' or noema correlated with the act"[27]; that is, "that the representational content is realized as an abstract entity -- the noema...."[28]

As for the case of the contents of perceptual states, however, Dreyfus tells us that (note again the characterization of perception as a skill) "... unlike Husserl's conceptualized noema, skills are not ideal, abstractable meanings. They cannot be entertained apart from some particular activation.... "[29] But nonetheless, these perceptual states play a fundamental role of meaning; as he goes on to say, "... these perceptual skills, like noemata, are the means through which we refer to and unify the objects of experience...."[30] So perception, like noemata, is seen by Dreyfus as playing exactly the kind of role that contentful states are to play -- means of reference, synthesizers of presentations (e.g. moments of time, slightly different perspectives on objects) -- but is not, as noemata are, a matter of "ideal, abstractable meanings".

Now the suggestion that "perceptual skills are not ideal, abstractable meanings" -- or to put in one of Dreyfus' more understandable ways, that acts of perception don't have abstractable meanings -- is most often tied up with the sorts of considerations which I discussed in the immediately preceeding section of this paper. The usual line of argument thus goes, "perception isn't a matter of rules and representations; hence perceptual states don't have abstract meanings." But if I'm right in claiming that Dreyfus doesn't give any good reason for thinking that perception isn't representational, or any for thinking that

either Husserl or a cognitivist must think that, then what reason is there to believe Dreyfus here? The only sort of independent consideration offered seems to be the comments, like the one in the quoted passage above, that perceptual states "cannot be entertained apart of some particular activation." Indeed, in a passage like this one, this latter point seems almost to be offered as a gloss on the idea that perceptual states don't have "abstractable meanings". The question is, even if it's true, why should this matter?

There are, as I see it, two ways to understand the the claim that perceptual states cannot be entertained apart from some particular activation. One is as the fairly straightforward idea that there are mental states which we can only as a matter of fact be put in by certain kinds of stimulations of our sensory transducers, or perhaps the afferent nerves from those transducers. And although it's not obvious what we should say about, say, <u>hallucination</u> here, there's surely something clearly right about the claim when it's taken in this way. Right, but pretty mundane. The sorts of mental presentations you can typically, as it were, generate at will (by, for example, imagination), are different from those typically generated by perception -- at least in terms of vividness, inescapability, and so on. But so what? Two points: First, Why should this be thought of as a difference in <u>content</u> of the states? And second, even if you think it should be thought of in that way (e.g.

you think that the kind of response given to the problem of meaning holism earlier requires it), why should that bother us? Once again, it might seem to pose a problem for a Husserlian view about the role of reflection in the consideration of mental contents, in that there could be contents which could not be entertained (i.e. representations which could not be tokened) just by sitting in the armchair and engaging in phenomenological investigation. But even if that's right, it doesn't seem to pose any special kind of puzzle for the view which is (as Dryfus acknowedges -- see above) both more central to Husserl's account of mental activity, and shared with the cognitivist -- view of mental activity as absract and rule-governed. After all, as Dreyfus at one point admits,

> ...Husserl himself suggests that in doing phenomenological psychology we could as well consult a test subject as consult ourselves, and it seems equally possible that we could just as well hypothesize the elements and structures or deduce them from overt behavior.[31]

The second way in which to understand the "no entertaining perceptual states apart from a particular activation" assertion is slightly different, and, I think, much more interesting. Here, the idea is to take this as a claim about the essentially non-solipsistic character of the semantic properties of perceptual states. On this way of understanding the point, the special connection between a perceptual state and, as Dreyfus likes to say, its

"conditions of satisfaction"[32], is not just, as in the suggestion above, the contingent one of those conditions in the world being (normally) the only way to get that state to occur. Rather, the connection is more intimate. It is, roughly, that what makes the perceptual state the one that it is -- what determines its content or its contribution to the content of our mental states -- is is the fact that it presents certain real external objects (or perspectival presentations of objects) in the world. Or to put it slightly differently: The contribution of perception to the (opaquely individuated, or narrow) contents of our mental states is not just a matter of the conceptual role of those perceptual states with respect to our other mental representations, but also depends on the relationships of those perceptual states to things which don't survive bracketing -- states of the external world. Perceptual states, on this reading, can't be separated from their particular activations in the sense that they can't be viewed as having the representational contents they do if you abstract from what states of the world they are actually semantically directed at. There is, as it were, no fully opaque reading of their contents. As Dreyfus says in criticizing Husserl at one point:

> ...it is only one step -- albeit a very dubious one -- from normal logical reflection directed toward the ideal correlates of referentially opaque conceptual acts to a special kind

> of reflection, the phenomenological reduction, in which Husserl claims to abstract the meanings of the referentially transparent acts of perception as well.[33]

Now I in fact believe that the central point here is correct. In fact, I have elsewhere -- in "The Chemistry of Intrinsic Intentionality", particularly section IV -- gone to some length in trying to make this point clearly and persuasively. So, rather than repeating such arguments here, let me instead accept the point from present purposes, and see what this suggests in the present context.

VI. Three grades of Semantic Involvement

So there would seem to be components of content tied with perceptual states which don't survive bracketing. The questions to ask then are (1) to what degree does this conflict with Husserl's line on intentionality, and (2) how does this bear on the parallel move of methodological solipsism in cognitive science? So first, let me re-ask the question: What role does the noema perform with repect to meaning? Husserl often says that it's only the noema that matters for intentionality, and Dreyfus takes this to mean that the noema is all there is to "representational content" or "meaning". Dreyfus' line:

> The noema, as conceived by Husserl, is a complex entity that has a difficult -- perhaps impossibly difficult -- job to perform. It must account for the mind's directedness towards objects. Therefore it must contain three components. One

> component must pick out a particular object outside the mind, another component must provide a "description" of that object under some aspect, and a third component must add a "description" of the other aspects which the object picked out could exhibit and still be the same object. In short, the noema must "refer", "describe", and "synthesize."[34]

This is an important passage. First and foremost, it needs to be emphasized that the noema alone needn't guarantee of any particular object outside the mind that it be picked out. That is to say, Husserl is not giving a theory of reference, or an account of de re attitudes. The noema is that part of what's phenomenologically accessible which is relevant to the fixation of reference. Nothing else "within" conciousness is relevant. The question is whether that means that nothing else is relevant at all. And the answer (Husserl's and the right one) is of course "no". The referent or intentional object of an intentional state is, on Husserl's view, not any mystical entity, but rather the real (typically physical) object at which it's directed. As he says, "I perceive the thing, the object of nature, the tree there in the garden; that and nothing else is the real object of the perceiving 'intention.' A second immanent tree... is nowise given...."[35] Or as Dreyfus puts it, "For Husserl... [an] act successfully refers, however, only if there is in fact an object with properties exactly as intended.[36]

A quick digression: Husserl sometimes has been

held to have been trying to give a theory of, as it were, de re attitudes, but the reasons for ascribing this to him are bad ones. Let me breifly mention them; they are, I think, of three kinds: The first is on the basis of an idealist (or even phenomenalist) reading of his metaphysical views. Now I'm inclined to think that this is the wrong way to interpret the metaphysical implications of Husserl's work, but I won't argue the point here. For present purposes, let me just point out that if Husserl's theory of intentionality depends fundamentally on some kind of phenomenalist metaphysics, the interest in it as a precursor to cognitive science diminishes considerably. What we were looking for was hints to an account of how intentionality is related to the material world, not how the material world is "created" via intentional states.

The second reason for ascribing to Husserl the attempt to give an acount of de re attitudes comes from running together his pre-phenomenological reduction views (primarily in the Logical Investigations) with the views he held during his "pure phenomenology" period, to which Ideas was central. In the earlier work, before Husserl adopted the "bracketing" approach, he did concern himself in part with giving a account of, for example, demonstrative reference. But the fact that he was concerned with such an account in the period before he adopted the view that bracketing was central to the methodology of the science of the mind certainly doesn't itself show that he

was still trying to give such an account -- or that he thought such an account was possible -- once bracketing had been adopted.

This kind of reason is sometimes conjoined with the third sort of reason for this view, which is based on a reading of what Husserl says about about the notion of the "determinable X" in the noema, primarily in chapter 11 of Ideas, "Noematic Meaning and Relation to the Object." A good example of this is to be found in David Smith and Ronald McIntyre's work on this; both in their book Husserl and Intentionality, and, perhaps even more clearly, in McIntyre's article "Intending and Referring." The claim there is that Husserl wants the "determinable X" of the noema to be, as McIntyre says, "correlated with the object itself" (i.e. the referent); and that the X is a "'non-descriptive' component of sense... which presents an act's object directly."[37] Now I won't go into this in detail, as it's slightly ouside the scope of the present discussion, but I think it is worth pointing out here that what Husserl is concerned with here is once again not relation to the referent of an act, but something like a consideration of the logical form of judgements. The "determinable X" is to capture the notion of "having a partiicualr one in mind" rather than that of a de re attitude.[38] There are, I believe, clear textual considerations in favor of taking Husserl this way (e.g. the fact that in this context he always puts 'object'

in quotes, his standard device for signifying that he is using a term with its post-bracketing, "altered" meaning [39], and his admonishment in the middle of the present discussion that "it must not be forgotton that all our discussions, including the ones now before us, are to be understood in the sense of the phenomenologicl reductions...."[40]). And in any case, it would seem that this reading is suggested by the guidlines of rational reconstruction; the is work is, at least in the present context, more interesting and relevant if taken in this way.

However, I said this wasn't a paper on Husserl, so let me get back to the point at hand. The point is, given the way I'm taking Husserl, that at least some semantic properties -- reference, and of course truth value -- are <u>not</u> just a matter of the noema itself. The noema may be the vehicle of reference, but it needn't be the noema alone which determines for itself a particular real object. Considerations involving the actual state of the external world -- like what objects it contains, and which I perceptually interact with -- may also enter into the determination of what (if any) particular object outside the mind is picked out. This is, of course, a contemptously familiar point from recent discussions of indexicality in the philosophy of language.[41]

So there is one of our "three grades of semantic involvement" which the noema is seemingly not intended to capture: that of reference and truth value (or in the

fashionable lingo, "wide content"). But there are, as noted above, differences in "narrow content" which are a matter not simply of relations between representations, but a matter of the relationships of perceptual representations to the external world. To take a simple example (which I have tried to exploit elsewhere): the narrow contents of color terms -- their meanings as "opaquely individuated" for use in the explanation of behavior -- depend essentially on what properties in the world they are perceptually related to in the right way. But this perceptual relationship is exactly the sort of thing which will not survive bracketing. So, bracketing doesn't give narrow content. Two questions: (1) Does this mean that the use of the notion of narrow content in cognitive science is out, because it isn't captured by this way of understanding the requirement of "methodological solipsism"? And (2) does this rule out the level of noema or formal rules alone as a reasonable level of consideration of intentionality? The answers here are, I think, "no" and "no". Here's why.

Question (1): Recall that the critical motivation for the solipsistic move in cognitive science was so that we wouldn't have to "wait forever" for a completed science before giving an account of intentional states. The problem was that semantic properties like reference could depend on the outcome of, say, chemistry. But it needn't be the case that the semantic features of perceptual states which enter into an evaluation of the narrow contents of mental

states depend on how all of science turns out. They may, for example, depend only on things like the laws of optics and a theory of transduction. If the way in which perception enters into narrow content is thought of not as diectly presenting referents, but as presenting something like observational presentations of objects, a theory of the "hidden essences" of objects may not be at all relevant. You wouldn't need a theory of the connection of thoughts to their referents, but one of the connection of perceptual states to the observationally salient properties of objects. Now perhaps there's no such theory, or no usable notion of "observationally salient properties of objects." But on the other hand, maybe there is.[42] Let's see if somebody comes up with one. If there is one, then although the notion of content in cognitive science may not be "presuppositonless" in the sense discussed earlier, at least it doesn't presuppose everthing -- just an ideal theory of psychophysics.

Question (2): If this is right, what's the other (non-psychophysical) component of a theory of narrow content? An account of conceptual roles, of course -- i.e. a story about the noematic structures. Perhaps the notion of noematic content doesn't capture narrow content on the whole, but it may well be exactly the sort of thing to be looked at in considering that aspect of narrow content which is independent of external considerations. From this level of analysis, the aspect of the contents of perceptual

representations dependent on their psychophisical ties with objects would be bracketed -- not denied, but simply excluded from the theory of computational activity.

Time to stop and draw the moral. The moral is that if this way of looking at things has any plausibility, we are then once again seeing the general notion of meaning broken down into distinct (and hopefully more precise) parts. There is not just one notion of meaning here, or even two ("wide" and "narrow"), but now three. I'm not sure which road this should suggest that we're on: the one to a more scientifically precise and clarified notion (or class of notions), or the one to the utter breakdown and eventual rejection of the entire class of notions. Perhaps it should be taken to suggest we're at the crossroads. Too bad the street signs aren't up yet.

NOTES

[1] Dreyfus (1982a), pp. 11-2.

[2] Dreyfus (1982a), p. 1.

[3] Husserl (1962), p. 100.

[4] Smith and McIntyre (1982), p. 95.

[5] Dreyfus (1982a), p. 14.

[6] Fodor (1980), p.225.

[7] Fodor (1980), p. 227.

[8] Fodor (1980), p. 231.

[9] Fodor (1980), p. 227.

[10] Dreyfus (1982a), pp. 14-5.

[11] Dreyfus (1982a) p. 3.

[12] Husserl, p. 100.

[13] Fodor (1980), p. 247.

[14] Fodor (1980), P. 248.

[15] See part 2 of this thesis, "Meaning Psychologized," for a discusssion of the meaning holism problem for conceptual role semantics.

[16] Dreyfus (1982a), p. 20.

[17] Dreyfus (1980), p. 78.

[18] Dreyfus (1982a), p. 17.

[19] Dreyfus (1982a), p. 23.

[20] Dreyfus (1982a), p. 23.

[21] Dreyfus (1982a), p. 21. There's an obvious parallel between this "what we know / what we are" distinction and the kind of distinction which Chomsky (1980) makes between "knowing that" and "knowing how". The obvious point: on Chomsky's way of making the distinction, there's no reason to think that knowing how isn't to be explained in terms of the "autonomous" mental realm of computation and representation.

[22] For a consideration of how this might in fact be the case, in particular in the case of mental imagery, see Block (1983).

[23] Dreyfus (1982a), p. 25.

[24] See Fitts and Posner (1967) for a nice discussion of this kind of skill learning from withing the cognitivistic framework.

[25] Dreyfus (among others) sometimes looks as though he wants to place some weight on the differences between conscious and unconscious processing -- for example SPEED -- in arguing that such unconscious processing is

non-representational. For an atractive alternative story about these differences, see Fodor (1983).

[26] Dreyfus (1982b), p. 108.

[27] Dreyfus (1982a), p. 7.

[28] Dreyfus (1982a), p. 9.

[29] Dreyfus (1982b), p. 122.

[30] Dreyfus (1982b), p. 122.

[31] Dreyfus (1982a), p. 14.

[32] See, e.g., Dreyfus (1980), p. 79.

[33] Dreyfus (1982b), p. 108.

[34] Dreyfus (1982a), p. 7.

[35] Husserl, p.243.

[36] Dreyfus (1982a), p. 5.

[37] McIntyre (1982), p. 227.

[38] For a nice discussion of this distinction, see Dennett (1982).

[39] Cf. Husserl, section 89.

[40] Husserl, p. 346.

[41] For my own story here, see part 2 of this thesis,

"Meaning Psychologized."

[42] See Fodor (unpublished-c) for a discussion of this.

REFERENCES

Block, Ned (1981a) "Psychologism and Behaviorism." Philosopical Review 90, pp. 5-43.

Block, Ned (1981b) "Troubles With Functionalism." In Readings in the Philosophy of Psychology, Vol. 1, Ned Block, ed. Cambridge, Mass.: Harvard University Press, pp. 268-305.

Block, Ned (1983) "Mental Pictures and Cognitive Science." Philosopical Review 92, pp. 455-512.

Burge, Tyler (1982) "Other Bodies." In Thought and Object: Essays on Intentionality, Andrew Woodfield, ed. Oxford: Clarendon Press, pp. 97-120.

Chomsky, Noam (1980) "Rules and Respresentations." Behavioral and Brain Sciences 3, pp. 1-15.

Dennett, D.C. (1982) "Beyond Belief." In Thought and Object: Essays on Intentionality, Andrew Woodfield, ed. Oxford: Clarendon Press, pp. 1-95.

Dretske, Fred (1981) Knowledge and the Flow of Information. Cambridge, Mass.: M.I.T. Press/Bradford Books.

Dreyfus, Hubert L. (1980) "Dasein's Revenge: Methodological Solipsism as an Unsuccessful Escape strategy in Psychology." Commentary on Fodor (1980a), Behavioral and Brain Sciences 3, pp. 78-9.

Dreyfus, Hubert L. (1982a) Introduction to Husserl, Intentionality, and Cognitive Science, Hubert L. Dreyfus, ed. Cambridge, Mass.: M.I.T. Press/Bradford Books, pp. 1-27.

Dreyfus, Hubert L. (1982b) "Husserl's Perceptual Noema." In Husserl, Intentionality, and Cognitive Science, Hubert L. Dreyfus, ed. Cambridge, Mass.: M.I.T. Press/Bradford Books, pp. 97-124.

Field, Hartry (1977) "Logic, Meaning, and Conceptual Role." Journal of Philosophy 74, pp. 379-409.

Fitts, P.M., and M.I. Posner (1967) Human Performance. Belmont, Calif.: Brooks/Cole.

Fodor, Jerry A. (1980a) "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." Behavioral and Brain Sciences 3, pp.

63-73.

Fodor, Jerry A. (1980b) "Searle on What Only Brains Can Do." Commentary on Searle (1980), Behavioral and Brain Sciences 3, pp. 431-2.

Fodor, Jerry A. (1983) The Modularity of Mind. Cambridge, Mass.: M.I.T. Press/Bradford Books.

Fodor, Jerry A. (forthcoming) "Semantics, Wisconsin Style." Forthcoming in Synthese.

Fodor, Jerry A. (unpublished-a) "Narrow Content and Meaning Holism." Unpublished paper.

Fodor, Jerry A. (unpublished-b) "Psychosemantics: Or, Where Do Truth Conditions Come From?" Unpublished paper.

Fodor, Jerry A. (unbublished-c) "Observation Reconsidered." Unpublished paper.

Harman, Gilbert (unpublished) "Conceptual Role Semantics."

Haugeland, John (1981) "Semantic Engines: An Introduction to Mind Design." In Mind Design, John Haugeland, ed. Mongomery, Vermont: Bradford Books, pp. 1-34.

Husserl, Edmund (1962) Ideas: General Introduction to Pure Phenomenology. Translated by W. R. Boyce Gibson. New York, New York: Collier Books.

Kaplan, David (unpublished) "Demonstratives." Unpublished manuscript.

Leibniz, Gottfried Wilhelm (1961) The Monadology. In Philosophical Classics, Vol. 2: Bacon to Kant, Walter Kaufmann, ed. Englewood Cliffs, New Jersy: Prentice Hall, pp. 205-14.

Lycan, William G. (1980) "The Functionalist Reply (Ohio State)." Commentary on Searle (1980), Behavioral and Brain Sciences 3, pp. 434-5.

McIntyre, Ronald (1982) "Intending and Referring." In Husserl, Intentionality, and Cognitive Science, Hubert L. Dreyfus, ed. Cambridge, Mass.: M.I.T. Press/Bradford Books, pp. 215-32.

Nagel, Thomas (1974) "What Is It Like to Be a Bat?" Philosophical Review 83, pp. 435-50.

Putnam, Hilary (1975) "The Meaning of Meaning." In Mind, Language and Reality: Philosopical Papers, Volume 2, by Hilary Putnam. Cambridge: Cambridge University

Press, pp. 215-71.

Putnam, Hilary (1981) Reason, Truth and History. Cambridge: Cambridge University Press.

Putnam, Hilary (1984) "Computational Psychology and Interpretation Theory."

Rey, Georges (1980) "The Formal and the Opaque." Commentary on Fodor (1980a), Behavioral and Brain Sciences 3, pp. 90-2.

Sartre, Jean-Paul (1957) The Transcendence of the Ego: An Existentialist Theory of Consciousness. Translated by Forrest Williams and Robert Kirkpatrick. New York, New York: Noonday Press.

Searle, John (1979a) "What Is an Intentional State?" Mind 88, pp. 74-92.

Searle, John (1979b) "Intentionality and the Use of Language." In Meaning and Use, A. Margalit, ed. Dordrecht, Holland: D. Reidel, pp. 181-197.

Searle, John (1980) "Minds, Brains, and Programs." Behavioral and Brain Sciences 3, pp. 417-424, 450-457.

Searle, John (1983) Intentionality: An Essay in the Philosophy of Mind. Cambridge: Cambridge University Press.

Smith, David Woodruff, and Ronald McIntyre (1982) Husserl and Intentionality: A Study of Mind, Meaning, and Language. Dordrecht, Holland: D. Reidel.

Stitch, Stephen P. (1982) "On the Ascription of Content." In Thought and Object: Essays on Intentionality, Andrew Woodfield, ed. Oxford: Clarendon Press, pp. 153-206.

Stitch, Stephen P. (1983) Folk Psychology and Cognitive Science: The Case Against Belief. Cambridge, Mass.: M.I.T. Press/Bradford Books.