A SHOCKING IDEA ABOUT MEANING

Michael DEVITT

1. Externalism

Is there any philosopher on Earth who has not heard about Twin Earth? I assume not and so will spare you the story. The conclusion that Hilary Putnam draws from Twin Earth is that the intrinsic "narrow" psychological states of Oscar on Earth and Twin Oscar on Twin Earth are insufficient to determine the reference of 'water'. For these states are the same and yet Oscar and Twin Oscar refer to different stuff, Oscar to H_2O, Twin Oscar to XYZ. As Putnam puts it in "The Meaning of 'Meaning'," "meanings just ain't in the head" (1975: 227). Later, in *Reason, Truth and History* (1981), using the case of a brain in a vat, he drew out the general implications of his discussion, pointing out that the brain's system of representation has no "*intrinsic*, built-in, magical connection with what it represents" (p. 5). Wittgenstein asked: "What makes my image of him into an image of *him*?" (1953: 177). Putnam's answers that nothing in the head is sufficient to do so.

This has an important consequence for the most popular theory of reference (until recently, at least), the description theory. The consequence is that a theory of this sort is *essentially incomplete*. A description theory explains the referential properties of a word by appealing to its associations with other words; a likely example is that 'bachelor' is explained by its associations with 'adult', 'unmarried', and 'male'. Suppose that this is right and so these associations with 'bachelor' do indeed explain the reference. Still, this could not be the *complete* explanation because the associations are entirely in the head of a speaker and Putnam's discussion shows that what is in the head is not sufficient to determine reference.

© Revue Internationale de Philosophie 4/2001 - n° 218 - pp. 471-494.

I am fond of another argument for the incompleteness of description theories. ([1]) How does a word's associations with other words explain its reference? It's reference is explained in that it refers to what all of the other words, or most of them, jointly refer to. Thus, in our example, 'bachelor' refers to whatever 'adult', 'unmarried', and 'male' jointly refer to. This immediately raises a question: What determines what these other words refer to? To complete the explanation of the reference of 'bachelor' we need to answer this question. Perhaps description theories can be used again. But this process cannot go on for ever: there must be some basic words whose referential properties are not parasitic on others. Otherwise, language as a whole is cut loose from the world. Description theories pass the referential buck. The buck must stop somewhere with a different sort of theory.

Putnam's discussion not only brings out the incompleteness dramatically, it shows us where we have to look for the ultimate explanation of reference: we must look to what is outside the head, presumably to some sort of causal relation between the speaker and the external world.

We should not suppose that all enthusiasts for description theories would deny this point. Indeed, the point may be implicitly acknowledged by those theorists who allow that a person who cannot describe the bearer of a proper name can nonetheless be competent with the name provided she can *recognize* the bearer; provided she can pick out the bearer in a lineup, saying "that person." Doubtless some of these theorists would agree that this sort of demonstrative reference is to be explained causally not descriptively. But one theorist at least would not: John Searle. ([2]) I suspect that he speaks for many.

Although description theories may well be right for many words, their essential incompleteness shows that they cannot be right for all. Another conclusion we should draw from Twin Earth is that description theories are not right for natural kind words like 'water' in particular. For, Oscar and Twin Oscar associate exactly the same

(1) Devitt and Sterelny 1987: 51-2, 70. Jackson (1998) has a neat response to this argument taking it to be concerned simply with the reference of "public" words. I think the argument should be taken as concerned primarily with the reference of parts of thoughts.

(2) His 1983 is a characteristically vigorous argument for the view that meanings *are* in the head. He proposes what is, in effect, a description theory of demonstrative reference. My 1990 is a defense of the Putnam line.

descriptions with 'water'. Yet they refer to different stuff. So the association of a description with the word is not sufficient to secure reference. Adapting the fantasy a little, its easy to see that description theories are not right for names either. Oscar and Twin Oscar associate the same descriptions with 'Clinton' and yet refer to different people. ([3])

Of course, Putnam provided other good reasons for rejecting the description theory of natural kind words. So did Saul Kripke (1980). And Kripke and Keith Donnellan (1972) provided other good reasons for rejecting the description theory of names. In my view, the most convincing of these reasons is what I call "the argument from ignorance and error": the beliefs that speakers have about things that they refer to are often too meager or too wrong to meet the demands of a description theory.

In sum, Putnam has the thesis:

EXTERNALISM
Some words, including names and natural kind words, refer in virtue of causal relations that are partly external to the head (and hence these words are to be explained by a theory that is not a description theory).

What causal relation explains reference? Putnam has a theory about this too, an historical-causal theory. ([4]) According to this theory, a natural kind word's reference is determined by a certain sort of causal link to samples of the kind, a link that starts from groundings in the sample and includes reference borrowings as the word is used in com-

(3) This refutation of description theories of natural kind words and names needs a qualification (as discussions with Kim Sterelny brought home to me). The refutation is based on the assumption that a description that both Oscar and Twin Oscar associate with a word *has the same referent*. For, if the description referred to one thing when in Oscar's head and another thing when in Twin Oscar's head, then that would explain why the word whose reference depends on that description has different referents for Oscar and Twin Oscar. And an associated description would have different referents for Oscar and Twin Oscar if it involved, implicitly or explicitly, indexical reference - for example, the description, 'the woman I most admire' — because the objects available for "demonstration" on Earth and Twin Earth are numerically different. So a description theory that required the reference-determining descriptions to be of this indexical sort would survive the Twin-Earth discussion. So far as I know, nobody has ever proposed such a description theory of names or natural kind words.

(4) Sadly he does not view his theory naturalistically. I have discussed this elsewhere: 1997a, particularly, pp. 330-8; 1997c.

munication ("the linguistic division of labor"; Putnam 1975: 227-8). In the case of a name, the grounding must be in the name's bearer. (5)

There is plenty of room for argument about the historical-causal theory (Devitt and Sterelny 1987: chs 4-5). And there are other externalist theories available, indicator and teleological theories. Perhaps some combination of these three sorts of theory will turn out to be right. It very much remains to be seen. My early discussion in this paper makes no assumption about which externalist theory is right. However, the later discussion, in sections 5 and 6, is not so neutral. It relies on one part of the historical-causal theory:

REFERENCE BORROWING
The reference of some words, including names and natural kind words, can be borrowed.

However, this lack of neutrality is not worrying because this part of the historical-causal theory should be adopted by any externalist theory. Putnam and others have made it seem very plausible that the reference of many words can be borrowed. Indeed, it is largely *because* the reference of a word can be borrowed that people competent with the word can be so ignorant and wrong about the referent. The part of the historical-causal theory that seems suspect is the theory of how reference is fixed by causal groundings. It is here, in the theory of reference *fixing*, that an indicator or teleological theory is likely to make a valuable contribution.

In *Coming to Our Senses* (1996), I argue for an idea about meaning that is partly based on EXTERNALISM. My aim in this paper (6) is to show how shocking this idea is. It is shocking because it goes against what I can't resist calling "four dogmas of semantics": (i) "the direct-reference dogma" that the meaning of a nondescriptive word like a name is simply its property of referring to whatever, (7) discussed in section 3; (ii) "the Cartesian dogma" that competence with a word involves (tacit) knowledge about its meaning, discussed in section 4; (iii) "the narrow dogma" that narrow meanings are needed for the scientifically proper explanation of behavior, discussed in section 5. (iv) "the rich dogma" that a cognitively rich and fine-grained meaning is needed to explain behavior, discussed in section 6. I doubt that anyone holds all four of these dogmas, but almost everyone holds at least one of them.

(5) Putnam's causal theory is to be found in "Explanation and Reference" and "Language and Reality" as well as in "The Meaning of 'Meaning'", all in his 1975. In the latter paper he argues that the theory even covers artifactual words like 'pencil' and social words like 'pediatrician' (1975: 242-5). I think that this goes too far (Devitt and Sterelny 1987: 5.4). It is important to the plausibility of the causal theory that reference is determined not simply by a dubbing but also by later groundings: typically a term is "multiply grounded" (Devitt 1981: 2.8, 5.4).

(6) The paper draws not only on my 1996 but also on 1997b and 1997d.

(7) More usually, it is the view that the meaning simply is the referent. But this is only a verbal difference.

I say that this shocking idea is partly based on Putnamian EXTERNALISM. It is also partly based on a Fregean thesis that I shall discuss in the next section.

In presenting Putnam's argument I have talked of the meaning and reference of "words." This naturally suggests that the concern is with linguistic items. Yet I think that the argument should be seen as primarily concerned with the mental states that linguistic items express: Putnam's point is that nothing in the head is sufficient to determine that *parts of thoughts* refer to water, Clinton, or whatever. And that is what I shall be primarily concerned with in what follows.

What are we to call these parts of thoughts? I believe in the language-of-thought hypothesis and so I shall call them "words." Still nothing hangs on this usage nor on the controversial language-of-thought hypothesis. The shocking idea applies to the parts of thoughts whatever those parts are.

2. The Shocking Idea

In claiming that "meanings just ain't in the head" Putnam has in mind a notion of meaning that determines reference. His argument shows that nothing in the head is sufficient to determine reference. So it is not sufficient to constitute a meaning that determines reference.

Is there anything more to the meaning of a word than its property of referring to whatever? Putnam thinks so, arguing for a multi-factor theory which includes a "stereotype" in the meaning. Two-factor theories are popular. They supplement a referential factor with a functional- (or conceptual-) role factor. A few philosophers reject a referential view of meaning altogether. In contrast to all of these views, I urge the following Fregean thesis that is the other basis for my shocking idea:

MEANINGS AS MODES

The meaning of a word is its property of referring to something in a certain way, its mode of reference. ([8])

How are we to choose among this range of opinions about meaning? How should we argue for a theory of meaning?

In *Coming to Our Senses*, I address this methodological question and apply my answer to argue for MEANINGS AS MODES. Here, briefly, is the argument. We should start by considering what meanings are supposed to do. What theoretical purpose do we serve by ascribing meanings to thoughts? Probably quite a few, but I want to focus here on one that is particularly important: a thought causes intentional behavior partly in virtue of its meaning. Next consider what the folk, rightly or wrongly, ascribe to a thought for that purpose of explaining behavior; what the folk are, in effect, treating as a meaning. Typically the folk attempt to explain behavior using what Quine calls an "opaque" thought ascription. The ascription involves an attitude verb like 'believes' and a 'that' clause. The 'that' clause ascribes the putative meaning. *What* putative meaning does an opaque 'that' clause ascribe? I argue that it ascribes a property *of referring to something in a certain way*; it ascribes a *mode of reference*. And that is *all* it ascribes. So unless a functional role or a stereotype plays a role in determining reference it is not part of this property.

Suppose that this argument is right. It does not settle what meanings are. It simply tells us what folk ascribe to explain behavior and hence what they are, in effect, treating as a meaning (on the assumption that explaining behavior is one theoretically interesting thing that meanings are supposed to do). It describes the semantic *status quo*. But perhaps the folk are wrong. Perhaps the property of a mental word that really does play a role in explaining behaviour, and hence is its meaning, involves a stereotype or a non-reference-determining functional role. Or perhaps the eliminativist is right and there are no thoughts: something else altogether explains behavior. Of course, if

(8) Actually I think (1996) that the semantic picture is a deal more complicated. Given our theoretical interest in meanings I think that we have good reason to believe that several properties of a word token are meanings. (This is not the familiar view that many word types have more than one meaning and so are ambiguous.) But these complications are beside the point of this paper and will be ignored.

the *status quo* is as we have described it, such positions are all revisionist. But that does not show that they are wrong.

Showing this is, of course, a lengthy business involving criticism of revisionism and its arguments. I shall do a little of this criticism later (secs 5 and 6). Now I want to note one simple, yet powerful, argument for the *status quo*. Briefly, it works. Day in and day out the folk use ordinary thought ascriptions to explain behaviour. For example, they say "Oscar believed that Mary was thirsty" to explain his giving water to Mary. And the folk are not alone in this habit: social scientists do it all the time too. Furthermore, these ascriptions appear to be, by and large, *successful*; the ascription to Oscar really does seem to explain his behaviour. This is evidence that thoughts really do have whatever properties the folk and social scientists ascribe to them, and that those properties really do explain behavior and so are meanings. And, I have argued, those properties are modes of reference. Given the explanatory success of this *status quo*, overthrowing it needs both a powerful argument and a plausible alternative semantics.

Putting the Fregean thesis, MEANINGS AS MODES, together with the Putnamian thesis, EXTERNALISM, yields the shocking idea. For, it follows from EXTERNALISM that some modes of reference, including those for names and natural kind words, are causal and non-descriptive. If the historical-causal theory is right for such a mode, it is a property of referring by a certain sort of causal chain. For example, the mode for 'Mark Twain' is the property of referring by means of causal chains grounded in Mark Twain and involving the sounds, inscriptions, and so on, that constitute the history of the name's use to designate Mark Twain; and the mode for 'Samuel Clemens' is similar but involves the sounds, inscriptions, and so on, of this different name. If another externalist theory is right for a word, its mode will be its property of referring by some other sort of causal relation to external reality. Put this together with MEANINGS AS MODES yields:

THE SHOCKING IDEA

The meanings of some words, including names and natural kind words, are causal modes of reference that are partly external to the head.

I emphasize that THE SHOCKING IDEA simply follows from the Putnamian thesis and the Fregean thesis and so cannot be denied without denying one of those theses.

The idea is shocking because it clashes with the four dogmas of semantics, which we will now consider in turn.

3. The Direct-Reference Dogma

The direct-reference dogma, (9) like THE SHOCKING IDEA, has its roots in the refutation of description theories for names and natural kind words. But according to direct reference, the meaning of such a nondescriptive word is simply its property of referring to whatever. (10) Indeed, the received view, even among those who do not subscribe to direct reference, is that the direct-reference view is a *consequence* of accepting nondescriptive causal theories of these words. (11) The combination of EXTERNALISM and MEANINGS AS MODES is not thought to be viable. Why not? The possibility that the meaning of a nondescriptive word might involve a nondescriptive causal mode of reference is either ignored, set aside, or dismissed as preposterous. Nathan Salmon is a striking example, describing an earlier proposal of mine as "ill conceived if not downright desperate… wildly bizarre… a confusion, on the order of a category mistake" (1986: 70-1). Yet he says almost nothing in support of this. What he most needs to provide is some principled way of judging what goes into the category of meaning.

The direct-reference philosopher thinks that the reference of a name is not determined descriptively. Yet he would agree that it must be determined somehow, presumably causally. So he cannot deny that a name *has* a causal mode of reference of some sort. What he needs to provide then is some basis for denying that this mode is the name's meaning. He identifies a name's meaning with its property of referring to its bearer. What is his basis for this identification rather than one with the name's property of referring to the bearer in a particular

(9) Some examples: Salmon 1981: 11; 1986; Barwise and Perry 1983: 165; Almog 1985: 615-6n; Wettstein 1986: 185, 192-4; Fodor 1987: 72-95; Crimmins and Perry 1989: 686; Braun 1991: 302.

(10) Where I use the technical term 'meaning' many prefer others; for example, 'propositional content' or 'semantic value'. Nothing hinges on this verbal difference.

(11) Some examples: Loar 1976; Ackerman 1979a: 58; 1979b: 6; McGinn 1982: 244; Baker 1982: 227; Lycan 1985; Block 1986: 660, 665; Lepore and Loewer 1986: 60; Wagner 1986: 452; Wettstein 1986: 187; Katz 1990: 31-4.

causal way? There is no doubt that the idea of a meaning as a causal mode of reference is alien to the semantic tradition, but that hardly counts as an adequate basis. The tradition may be wrong! I have argued that the meaning should be identified with the mode of reference because that is the property of a name that causes behavior. And causing behavior is one thing that meanings are supposed to do: it is one thing that makes meanings theoretically interesting. Direct-reference philosophers need to justify some other account of the role of meanings to support their view that the meaning is simply the role of referring. To my knowledge, no such justification has ever been attempted, let alone provided. In its absence, the direct reference dogma is left theoretically arbitrary and *ad hoc*.

4. The Cartesian Dogma

The Cartesian dogma is ubiquitous in philosophy and linguistics. It is the view that for a speaker to be able to use an expression with a certain meaning, or think a thought with a certain meaning, is for her to (tacitly) *know that* it has that meaning. This Cartesianism stems from the idea that, simply in virtue of having the mental state of being linguistically competent with an expression, a speaker is in a position to discover facts about it. These facts include facts about the meaning of the expression. So, simply in virtue of her competence, the speaker has some sort of "privileged access" to facts about meaning. To get knowledge of a meaning she does not have to carry out the sort of empirical investigation of the world that knowledge usually requires, she can simply "look inwards."

In linguistics we find Cartesianism in the standard view of the intuitive grammatical judgments of the competent speaker. The speaker is thought to derive these judgments by a causal-rational process from a representation of the grammar in her "language faculty." In philosophy Cartesianism seems to be an almost unquestioned part of the traditions of Frege and Russell. Consider these typical statements:

> It is an undeniable feature of the notion of meaning… that meaning is *transparent* in the sense that, if someone attaches a meaning to each of two words, he must know whether these meanings are the same. (Dummett 1978: 131)

> The natural view is that one has *some kind of* privileged semantic self-knowledge. (Loar 1987: 97)

Consider also the received Fregean view that '$a = a$' and '$a = b$' must differ in meaning *because* they differ in informativeness, perhaps the most important reason why many came to believe in MEANINGS AS MODES. Why would one suppose that a difference in informativeness, an epistemic difference, established a semantic difference unless one supposed that the speaker had access to the semantic facts? (12)

I have always presented the historical-causal theory of names and natural kind words in a way that emphasized its anti-Cartesianism. Competence with such a word is simply an ability with it that is gained in a grounding or reference borrowing. A person has the competence in virtue of being linked appropriately into the causal network for the word. Competence does not require any *knowledge about* the meaning, any *knowledge that* the meaning is the property of referring by a certain type of causal chain. Indeed, reflection on Putnam's discussion suggests that Cartesianism *has* to be abandoned for aspects of meaning constituted "outside the head", the aspects on which all meanings ultimately depend. I like to argue that Cartesianism should be abandoned even for aspects of meaning constituted "inside the head." Still, for such aspects, there might seem to be some hope of explaining how a speaker's competence can amount to knowledge about the meaning. Thus, assuming the description theory for 'bachelor', we might hope to explain how someone competent with 'bachelor' must know that it applies to whatever 'adult unmarried male' applies to. For, the association of 'bachelor' with 'adult unmarried male' is at least in the head. But insofar as the meaning of a word is not in the head, the Cartesian view of it seems hopeless. How could any amount of reflection on what competence alone puts in the head establish such a highly theoretical, empirical, and external fact as that the meaning of 'Clinton' is the particular way in which it is causally

(12) I agree, of course, that the identity statements do differ in meaning but I disagree with this reason for thinking they do. I argue that they differ in meaning because they play different roles in the explanation of behavior (1996: 4.7).

Two other examples of Cartesianism: (1) the view that semantic competence consists in knowledge of truth conditions (on this, see Heidelberger 1980); (2) The description theorist's view that competence with a name provides identifying *knowledge* of its bearer.

linked to the person who happens to be President of the United States? This fact is surely entirely beyond the ken of the ordinary speaker.

The idea that there is a clash between the historical-causal theory and Cartesianism is familiar enough. Critics of the theory were quick to point out the clash and to urge that, *for this very reason*, the theory could not be right. It was alleged that the theory provided no account of what it is to "grasp the meaning" of a word, no account of the sort of knowledge that this requires. Gareth Evans is typical. Considering the possibility of explaining causally the different contents of two coreferential states, he claims that it is

> quite obscure how... the sheer difference between the causal relations could generate a difference in *content* between the two mental states, given that it need not in any way impinge on the subject's awareness. (1982: 83) (13)

Given the clash between any externalist causal theory and Cartesianism, one of them has to go. It should be clear that Cartesianism is the one, shocking as this may seem. For, description theories cannot be true for all words. Some meanings have to be constituted from relations that are partly external to the head. Competence alone could not yield knowledge of these meanings.

Of course, THE SHOCKING IDEA does not rule out Cartesianism for the internal aspects of meaning. Still, even that much Cartesianism needs much more support than it has ever been given. What could the process be by which competence alone leads not only to the formation of a belief about meaning but also to its justification? I suggest that we have no idea of the answer to this question. I have argued (1981: 95-110) that we should settle for a more modest view of linguistic and conceptual competence in general: these competencies are simply skills, matters of knowledge how not knowledge that. But that is another matter.

(13) See also Wettstein 1986: 194; Cassam 1989. My 1985 is a defense of the historical-causal theory from Evans' criticisms.

5. The Narrow Dogma

The narrow dogma is known as "methodological solipsism":

> the conviction that the best explanation of behavior will include a theory invoking properties supervenient upon the organism's current, internal physical state. (Stich 1978: 576)
>
> beliefs play a role in the agent's psychology just in virtue of intrinsic properties of the implicated internal representations... those properties of representations that can be characterized without adverting to matters lying outside the agent's head. (McGinn 1982: 208)

This intuition is supported by the view that a person and her functional duplicate — for example, Oscar and Twin Oscar — must be psychologically the same. (14)

THE SHOCKING IDEA posits causal meanings that are partly external to the head ; they are not narrow but as wide as a barn door. Part of the argument for the IDEA was that these meanings explain behavior. So, put the the IDEA together with its argument and we seem to have a clear clash with the narrow dogma.

I am in no position to deny the appeal of the narrow dogma having once embraced it somewhat tentatively (1989). I now think that it is quite mistaken.

To assess it we need to distinguish two views of narrow meaning. According to one, (15) the narrow meaning of a sentence is a function taking an external context as argument to yield a wide meaning as value. So, on this view, narrow meanings partly determine reference; they are intentional wide meanings "minus a bit"; they are "proto-intentional." The belief that we need only these meanings to explain behavior is, therefore, only moderately revisionist. According to the other, more popular, view of narrow meaning, (16) that meaning is a functional role involving other sentences, proximal sensory inputs, and proximal behavioral outputs. These meanings are not reference determining and differ greatly from the meanings that we currently ascribe. The belief that we need only these meanings to explain behavior is, therefore, highly revisionist.

(14) See, for example, Stich 1978: 574; Loar 1983: 665.

(15) To be found, e.g., in White 1982; Fodor 1987: 44-53.

(16) To be found, e.g., in Loar 1981 and 1982, McGinn 1982, Block 1986. My 1989 muddled the two views of narrow meanings.

I think that the highly revisionist version of the dogma has little to be said for it, despite its popularity. But I shall set it aside until the next section. The moderately revisionist version is more tricky and, in my view, more interesting.

If we believe in wide referential meanings, as I do, then we should have no objection to the idea of narrow meanings as functions, because any theory that explains the wide meaning will explain the narrow ones; we get the narrow meanings by abstracting from the links to context that partly constitute the wide meanings. Suppose someone has a mental word 'water' that has a wide meaning involving reference to water. Then it follows that the word has a narrow meaning that is a function yielding that wide meaning as its value given something about the watery world as its argument. Given something else as its argument, the function might yield a wide meaning involving reference to Twin water or whatever.

So we should accept that there are these sorts of proto-intentional narrow meanings. What sort of behavior might these meanings explain? Not intentional behavior, for they do not "reach beyond the head." But presumably there are "proto-intentional behaviors," just as there are proto-intentional meanings: intuitively but roughly, they are intentional behaviors minus the details external to the behaver. (17) If so, we should accept that just as wide meanings explain intentional behaviors, narrow meanings (as functions) explain proto-intentional behaviors. So what remains of the clash between THE SHOCKING IDEA and this version of the narrow dogma? Not much. The issue comes down to whether or not a scientific psychology should explain intentional or proto-intentional behavior.

This is not a big issue but I think there are good reasons to think that the dogma is on the wrong side of it. What does the dogma rest on? Largely, I think, on intuition. But it does rest partly on Twin-Earth examples. These examples are thought to show that psychological interest is only in what is common to a person and her twin. So that is all that a psychological theory should be concerned with: it should be concerned with narrow explanations. Thus, the explanation, "Oscar

(17) If one subscribes to the view that intentional behavior A essentially involves the wide intention to A, then the proto-intentional "part" of A essentially involves the narrow proto-intention to A. That part will be found in any intentional behavior that *fulfils* the proto-intention.

gave water to Mary because he believed that Mary was thirsty," is thought to bring in irrelevancies: we want an explanation that would also apply to Twin Oscar giving Twin water to Twin Mary.

Twin-Earth examples are misleading. The wide meaning that explains intentional behavior can be broken down into an internal narrow meaning and an external context. If the narrow meaning does almost all the explanatory work here it might be appropriate to focus on it and hence change to an explanation of proto-intentional behavior. But if the external context does a lot of the explanatory work, we lack motivation for this change. Twin-Earth examples are misleading because Oscar and Twin Oscar are identical, and their behaviors are so similar, that we are not encouraged to think that these behaviors may be largely explained by the external contexts. Rather, we are encouraged to think that the different external contexts for Oscar and Twin Oscar are responsible for the relatively minor differences between Oscar's giving water to Mary and Twin Oscar's giving Twin water to Twin Mary, but each behavior is almost entirely explained by a fine-grained narrow meaning shared by Oscar and Twin Oscar. So Twin-Earth examples suggest that some fine-grained narrow meaning carries most of the burden of psychological explanation. If the doctrine REFERENCE BORROWING is correct, this suggestion is false.

REFERENCE BORROWING, it will be remembered, covers names and natural kind words and perhaps some other words as well. The key thing about it is the tiny demand it places on the beliefs and capacities of competent speakers. It allows speakers to be very ignorant and wrong about the referent. Consider the famous example of 'elm' and 'beech'. Putnam's head seems to contain no description, image, or recognitional capacity that would enable him to distinguish elms from beechs. A consequence of REFERENCE BORROWING is that someone competent with a word for some sort of tree need not have anything in her head that picks out that sort of tree rather than any other. Indeed, on an extreme version of the doctrine, there may be nothing in her head that picks out that sort of tree rather than some sort of animal, or whatever. *According to REFERENCE BORROWING, the reference of a person's word may be determined very little by what is in her own head but very much by how that head is related to a society of heads and by that society's relation to reality*. So not much of the word's mode of reference is determined by what is in each head, not much of the wide meaning is narrow.

Is there *anything* in Putnam's head that distinguishes the modes of reference for 'elm' and 'beech' and hence distinguishes their narrow meanings? Consider the difference between Putnam's borrowing 'elm' and borrowing 'beech' in a communication situation. When he borrows 'elm' he experiences some conventional physical form of the word, for example, the sound /elm/. When he borrows 'beech' he experiences a different physical form, for example, the sound /beech/. If REFERENCE BORROWING is correct, the mental effects of these different sorts of experience is all there is to the difference between the narrow meanings of 'elm' and 'beech'. The narrow meaning of the one links it to one set of formal properties, that of the other, to another. That's it. So suppose that there is a world, *W*, just like Earth, but where the conventions are different and beeches are called 'elm'. Then Putnam's 'elm' referring to elms on Earth shares a narrow meaning with *W*-Putnam's 'elm' referring to beeches on *W*. In light of this, we see that the difference in the narrow meanings of 'elm' and 'beech' on Earth is trivial.

We can capture the point as follows. The proto-intentional narrow meaning of a word covered by REFERENCE BORROWING is "promiscuous" and "coarse-grained." It is promiscuous in that it can yield any of a vast range of referential values by changing the relevant external context as argument. It is coarse-grained in that there is very little to it.

What does this discussion of narrow meaning show us about the explanation of behavior? To explain Oscar's intentional act of giving water to Mary we advert to the wide meaning of Oscar's thoughts. That wide meaning is the value yielded by a certain narrow meaning when given the Earthly context as argument. Now that very same narrow meaning, with appropriate external contexts as arguments, could yield wide meanings that explain a vast range of intentional behavior. That behavior could involve not just water and Twin water but mercury, gold, and perhaps plastic, gin, and all the other stuffs. It could involve not just Mary and Twin Mary but Clinton, France, and any other object that can be named. Indeed, if REFERENCE BORROWING stretches far enough, the behavior might involve not only giving but taking, kicking, and many other acts. Where REFERENCE BORROWING applies, *a great deal of what explains intentional behavior, is determined by what is outside the head. The contribution of narrow meaning to this explanation is small*. Putting this another

way, the proto-intentional behaviors that these narrow meanings can explain are so coarse-grained and promiscuous as not to distinguish behaviors involving any named object, any stuff, and so on. What Twin Earth encouraged us to think is far from the truth.

Putnam has made us all used to the idea that Oscar's twin on some other planet might use 'water' to refer to XYZ. But, according to REFERENCE BORROWING, he might use it to refer to any other natural stuff at all. And, why the restriction to natural stuff? Isn't the theory just as plausible for unnatural stuff like plastic and gin? So perhaps Oscar's twin on some other planet might refer to any stuff at all by 'water'! The narrow meaning of 'water' places only trivial constraints on what it might refer to in a different context. It and the behavior it explains are highly promiscuous.

Of course, our verdict only concerns words covered by REFERENCE BORROWING. Very likely, many words are not, particularly those explained by a description theory. The narrow meaning of these words may be partly constituted by associations with others that will constrain what they could refer to in a context. Thus, perhaps 'bachelor' could refer to something in a context only if 'unmarried' refers to it in that context; so its narrow meaning is finer-grained and less promiscuous than those covered by REFERENCE BORROWING. ([18])

In sum, when REFERENCE BORROWING is added to THE SHOCKING IDEA, we see that the narrow meanings of many words, and the proto-intentional behaviors that these meanings explain, are coarse-grained and promiscuous. One might respond to this by rejecting REFERENCE BORROWING but, as I have emphasized, this doctrine has been made very plausible, particularly for names and natural kind words.

Where does this discussion leave the clash that triggered it, that between THE SHOCKING IDEA and the moderately revisionist version of the narrow dogma? That clash came down to the not very big issue of whether or not a scientific psychology should explain intentional or proto-intentional behavior. I take it that the more coarse-grained and promiscuous the proto-intentional behavior, the less appropriate it is for psychology to prefer to explain that behavior. And we have seen that a lot of proto-intentional behavior may be very

(18) I think (1996: 40-2, 290-1n) this discussion supplies an answer to Block's adaptation (1991: 60-1; 1994-5) of Putnam's ingenious Ruritania example (1983: 144-7).

coarse-grained and promiscuous. At this time, we have surely little reason to prefer proto-intentional behavior.

Finally, how is it *possible* that the explanation of intentional behavior might be as external as our discussion suggests? The role of a thought in explaining intentional behavior depends on what it refers to. We cannot settle a priori the extent to which intrinsic properties of the head determine this. What matters to the explanation is simply that the thought has its referential properties under modes appropriate to the behavior it is supposed to explain. Theories of reference will tell us about those modes.

6. The Rich Dogma

The rich dogma is that only a cognitively rich and fine-grained meaning could explain behavior. Ken Taylor has expressed the view nicely. Impressed by the fact that people who share a belief can be very cognitively diverse, he thinks that the belief could not be adequate to explain their behavior. What they share is not "causally homogeneous enough to back causally deep explanations" (1997: 363); it "is too conceptually thin" (p. 368). ([19]) We must ascribe a much richer meaning to mental words. This does not lead Taylor into holism — the view that every aspect of a word's functional role goes into its meaning — but it has led many others.

The rich dogma is surely related to the narrow one. Thus the fine-grained meaning demanded by the rich dogma is sought in the brain. So it seems natural to think that what explains behavior must be internal to the brain, as demanded by the narrow dogma. Ned Block (1991) is an example of a philosopher who holds both dogmas.

THE SHOCKING IDEA posits causal meanings that are likely to be cognitively austere and as poor as a church mouse, as we have just seen with the help of REFERENCE BORROWING: very little of the mode of reference of the likes of 'water' is "in the head." And, according to our argument for the IDEA, these causal meanings explain behavior. In contrast, according to the rich dogma only a "descriptive" property constituted in part by many, if not all, of a word's inferential

(19) Bertolet (1997) holds to the rich dogma but, unlike Taylor, is not a revisionist because he believes that the folk do ascribe rich meanings.

connections to other words could do the explanatory job. Since these descriptive properties are very different from the meanings that we currently ascribe, the rich dogma is highly revisionist.

If the presence of a property explains a certain behavior in one case it should do so in another. The intuition sustaining the rich dogma is that having a cognitively austere belief could not do this explanatory job because the people who have it can be so cognitively different. Only a much more fine-grained property could explain a similarity of behavior. But *how much* finer grained must the property be? It is not surprising that this intuition leads many to holism. For, unless we go all the way to holism, there will still be some cognitive diversity among those who share the explanatory property. That is, there will be diversity unless we require that the explanatory property of the word is constituted by *all* of the word's functional relations, with the result that the property can explain the behavior only of functionally identical people.

Holism is hopeless in semantics as it is hopeless in general. ([20]) One reason for thinking this is that whatever our purposes — explanatory, practical, even frivolous — they are not served by ascribing holistic properties. Or so I have argued (1996: ch. 3). Yet the rich dogma seems to drive us into holism. There must be something wrong with the dogma. And there is.

Very likely, anything differs from anything else in some respect. Yet, anything is similar to anything else in some respect. Among the similarities between things we often find a property that explains the behavior or characteristics of those things. The things that can be explained by this property can otherwise differ. Perhaps sometimes they will differ a lot, perhaps sometimes a little. But the mere fact that they differ a lot is no reason for concern about the explanation. A penguin is very different from an eagle but the fact that they are both birds explains a lot about them; Clinton is very different from Joyner-Kersee but the fact that they are both Americans explains a lot about

(20) Confirmation holism is an exception, in my view, but it is a different sort of holism from the ones I am referring to. It is at the level of realization not of constitution. It is not about the nature of the property, **being confirmed**; indeed, the property may even be atomistic. It is about the realization of the property in a sentence: Whether or not the sentence has the property depends on its relations to all other sentences in the web of belief, and on whether *they* have the property.

them. A car wheel is very different from a Ferris wheel but the fact that they are both wheels explains a lot about them. There is no scientific principle that says: "If a property is to be explanatory, things that share it must only differ to degree *n*." No whistle is blown when explained things differ a lot. There is no way of telling a priori that so much difference is too much.

It is beside the scientific point that the people who have a mental word with the wide referential property we ascribe are cognitively diverse. What matters is whether ascribing such properties best explain behaviors. The fact that these ascriptions are so successful in ordinary life and the social sciences is evidence that they are good. Do they face rivals that are as good? We have just taken a somewhat dim view of one: a semantics that ascribes narrow proto-intentional properties. But the rival that rich dogmatists have in mind is nearly always a semantics that posits narrow meanings of the other sort, functional-role properties involving links to proximal stimuli, proximal behavior, and other sentences. This is also the favored semantics of the narrow dogmatists, set aside in the last section. What is to be said for it?

Very little, I think. First, if we accept that the holistic view of these meanings is hopeless, we need some principled basis for constructing the meanings out of some but not all of its functional roles. No such basis has ever been provided: the meanings are left unexplained. Second, we have been given no idea how such meanings *could* explain intentional behaviors (like giving water to Mary) and it seems very unlikely that they could. Third, if they do not explain these behaviors, then revisionism requires that intentional behaviors be denied altogether; for if there are these behaviors and they are not explained by these narrow meanings then we need some other meanings to explain them, presumably the familiar wide meanings of folk psychology. We have been given no reason to deny intentional behaviors. All this is bad news for the highly revisionist narrow and rich dogmas. They have a heavy onus arising from the apparently striking success of our present practice of ascribing wide and austere causal meanings to explain behavior. Why do these ascriptions seems so successful if they are not really? What reason have we for thinking that the ascriptions that would be recommended by this revisionism would do any better? The dogmas have hardly begun to discharge their onus.

It is of course possible that future psychology will show that

behavior is best explained by meanings that are narrow or rich or both. Many are clearly convinced that this is so. But I doubt that there is any evidence that it is so.

7. Conclusion

Combining Putnam's externalism with the Fregean thesis that meanings are modes of reference yields the idea that the meanings of some words are partly external causal modes of reference. This idea clashes shockingly with four widely held views: (i) the direct-reference dogma that the meaning of a nondescriptive word like a name is simply its property of referring to whatever; (ii) the Cartesian dogma that competence with word involves (tacit) knowledge about its meaning; (iii) the narrow dogma that narrow meanings are needed for the scientifically proper explanation of behavior; (iv) the rich dogma that a cognitively rich and fine-grained meaning is needed to explain behavior. Shocking as the idea is, I think it is true for reasons I have indicated here and given at greater length elsewhere (1996). ([21])

The City University of New York.

(21) A version of this paper was delivered in July 1998 at the annual conference of the Australasian Association of Philosophy at Macquarie University, in September 1998 at a conference on Putnam's work, "Swimming in XYZ," in Karlovy Vary, and in October 1998 at Stanford University. I am indebted to the comments it received.

References

Ackerman, Felicia (Diana). 1979a. "Proper Names, Propositional Attitudes and Non-Descriptive Connotations." *Philosophical Studies* 35: 55-69.

——. 1979b. "Proper Names, Essences and Intuitive Beliefs." *Theory and Decision* 11: 5-26.

Almog, Joseph. "Form and Content." *Nous* 19: 603-616.

Baker, Lynne Rudder. 1982. "Underprivileged Access." *Nous* 16: 227-242.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes*. Cambridge, MA: MIT Press.

Bertolet, Rod. 1997. "Meaning, Cognitive Significance, and the Causal Theory. In Jutronic 1997: 175-188.

Block, Ned. 1986. "Advertisement for a Semantics for Psychology." In French, Uehling, and Wettstein 1986: 615-678.

——. 1991. "What Narrow Content Is Not." In *Meaning in Mind: Fodor and His Critics*. Barry Loewer and Georges Rey, eds. Oxford: Basil Blackwell. 1991: 33-64.

——. 1994-5. "An Argument for Holism." *Proceedings of the Aristotelian Society* 95: 151-169.

Boolos, George, ed. 1990. *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge: Cambridge University Press.

Braun, David. 1991. "Proper Names, Cognitive Contents, and Beliefs." *Philosophical Studies* 62: 289-305.

Cassam, Quassim. 1989. Review of Devitt and Sterelny 1987. *Mind* 98: 313-315.

Crimmins, Mark, and John Perry. 1989. "The Prince and the Phone Booth: Reporting Puzzling Beliefs." *Journal of Philosophy* 86: 685-711.

Davidson, Donald, and Gilbert Harman, eds. 1972. *Semantics of Natural Language*. Dordrecht: D. Reidel.

Devitt, Michael. 1981. *Designation*. New York: Columbia University Press.

——. 1985. Critical notice of Evans 1982. *Australasian Journal of Philosophy* 63: 216-232.

——. 1989. "A Narrow Representational Theory of the Mind." In *Rerepresentation: Readings in the Philosophy of Psychological Representation*, Stuart Silvers, ed. Dordrecht: Kluwer Academic Publishers: 369-402.

——. 1990. "Meanings Just Ain't in the Head." In Boolos 1990: 79-104.

——. 1997a. *Realism and Truth*. 2nd edn with a new afterword. Princeton: Princeton University Press.

——. 1997b. "A Priori Convictions about Psychology: A Response to Sosa and Taylor." In Villanueva 1997: 371-385.

——. 1997c. "On Determining Reference." In *Sprache und Denken / Language and Thought*, Alex Burri, ed. New York: Walter de Gruyter: 112-121.

——. 1997d. "Responses to the Maribor Papers." In Jutronic 1997: 353-411.

——, and Kim STERELNY. 1987. *Language and Reality: An Introduction to the Philosophy of Language*. Oxford: Basil Blackwell. (2nd edn 1999.)

DONNELLAN, Keith S. 1972. "Proper Names and Identifying Descriptions." In Davidson and Harman 1972: 356-379.

DUMMETT, Michael. 1978. *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press.

EVANS, Gareth. 1982. *The Varieties of Reference*, ed. John McDowell. Oxford: Clarendon Press.

FODOR, Jerry A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

FRENCH, Peter A., Theodore E. UEHLING, Jr., and Howard K. WETTSTEIN, eds. 1980. *Midwest Studies in Philosophy Vol. 5, Studies in Epistemology*. Minneapolis: University of Minnesota Press.

——. 1986. *Midwest Studies in Philosophy Vol. 10, Studies in the Philosophy of Mind*. Minneapolis: University of Minnesota Press.

HEIDELBERGER, Herbert. 1980. "Understanding and Truth Conditions." In French, Uehling, and Wettstein 1980: 401-410.

JACKSON, Frank. 1998. "Reference and Description Revisited." *Philosophical Perspectives 12: Language, Mind, and Ontology, 1998*, James E. Tomberlin, ed. Oxford: Blackwell Publishers: 201-218.

JUTRONIC, Dunja, ed. 1997. *The Maribor Papers in Naturalized Semantics*. Maribor: Pedagoska fakulteta.

KATZ, Jerrold. 1990. "Has the Description Theory of Names Been Refuted?" In Boolos 1990: 31-61.

LEPORE, Ernest, and Barry LOEWER. 1986. "Solipsistic Semantics." In French, Uehling, and Wettstein 1986: 595-614.

LOAR, Brian. 1976. "The Semantics of Singular Terms." *Philosophical Studies* 30: 353-377.

——. 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.

——. 1982. "Conceptual Role and Truth-Conditions." *Notre Dame Journal of Formal Logic* 23: 272-283.

——. 1983. "Reply to Fodor and Harman." In *PSA 1982* Vol 2, Asquith, P. D., and T. Nickles, eds. East Lansing, MI: Philosophy of Science Association: 662-666.

——. 1987. "Subjective Intentionality." *Philosophical Topics* 15: 89-124.

LYCAN, William G. 1985. "The Paradox of Naming." In *Analytical Philosophy in Comparative Perspective*, B. K. Matilal and J. L. Shaw, eds. Dordrecht: D. Reidel: 81-102.

KRIPKE, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press. [This is a corrected version of an article of the same name (plus an appendix) in Davidson and Harman 1972, together with a new preface.]

MCGINN, Colin. 1982. "The Structure of Content". In *Thought and Object*, A. Woodfield, ed. Oxford: Clarendon Press: 207-258.

PUTNAM, Hilary. 1975. *Mind, Language and Reality: Philosophical Papers Vol. 2*. Cambridge: Cambridge University Press.

——. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

——. 1983. *Realism and Reason: Philosophical Papers Vol. 3*. Cambridge: Cambridge University Press.

SALMON, Nathan. 1981. *Reference and Essence*. Princeton: Princeton University Press.

——. 1986. *Frege's Puzzle*. Cambridge, MA: MIT Press.

SEARLE, John R.1983. *Intentionality: An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

STICH, Stephen P. 1978. "Autonomous Psychology and the Belief-Desire Thesis." *Monist* 61: 573-591.

TAYLOR, Kenneth A. 1997. "Same Believers." In Villanueva 1997: 357-369.

VILLANUEVA, Enrique, ed. 1997. *Truth: Philosophical Issues, 8, 1997.* Atascadero: Ridgeview Publishing Company.

WAGNER, Steven J. 1986. "California Semantics Meets the Great Fact." *Notre Dame Journal of Formal Logic* 27: 430-455.

WETTSTEIN, Howard. 1986. "Has Semantics Rested on a Mistake?" *Journal of Philosophy* 83: 185-209.

WHITE, Stephen L. 1982. "Partial Character and the Language of Thought." *Pacific Philosophical Quarterly* 63: 347-365.

WITTGENSTEIN, Ludwig. 1953. *Philosophical Investigations*. Trans. G. E. M. Anscombe. 2d ed. rev. 1958. Oxford: Basil Blackwell.