

## Meaning and Use

MICHAEL DEVITT  
The Graduate Center, City University of New York

Part I argues that the use theory in Horwich's *Meaning* does not give sufficient attention to the relation between language and thought. A development of the theory is proposed that gives explanatory priority to the mental. The paper also argues that Horwich's identification of a word's meaning by its role in explaining the cause of sentences should be broadened to include its role in explaining the linguistic and nonlinguistic behavior that sentences cause. Part II argues that Horwich greatly overstates the case for his use theory; that the arguments from ignorance and error against description theories of reference can be adapted against the use theory; and that a tempting development of the use theory would risk both the collapse of the theory into truth referentialism and the difficulties that have plagued truth referentialism. Finally, a consideration of our ordinary thought ascriptions provides evidence against any use theory.

### 1. Introduction

Paul Horwich's *Meaning* (1998b) is an important and timely contribution to the debate at the foundations of semantics, a debate about what sort of property a meaning is.

Probably the most popular view of meaning has been a *truth-referential* one according to which the meaning of a sentence is explained in terms of its truth conditions and the meaning of a word is explained in terms of its referential properties. To give truth and reference these explanatory roles is to suppose that they are substantial and significant properties. Thus truth is usually thought to consist in some correspondence in nature between a belief or statement and the world. In recent times, a few philosophers have rejected this tradition and urged *deflationary* theories of truth and reference. Horwich has taken the lead with his influential earlier book, *Truth* (1998a).

On the deflationary view, crudely, truth isn't anything.<sup>1</sup> In particular it isn't anything that could play the explanatory role required of it in a truth-referential semantics. If this is right, then we need some other account of meaning (unless we abandon meaning altogether). Horwich provides a sketch of such an account in *Meaning*. Inspired by Wittgenstein, he provides a "use

theory." Another deflationist, Robert Brandom, has recently done so also (1994). Brandom's account of meaning rests on unexplained normativity, as he frankly acknowledges. This feature of his account is, of course, very troubling to the naturalistically inclined. Horwich's theory does not have this feature. Indeed, he does not think that normativity is intrinsic to meaning (although facts about meaning, like many nonnormative facts, can have normative import).

Part of the appeal of a use theory of meaning comes from the appeal of a deflationary view of truth (and reference). And that appeal comes largely from the deflationary view of the truth *term*. Deflationists argue, very plausibly, that the role of this term can be accounted for without taking it as standing for some substantial explanatory property. We then need a non-truth-referential semantics. The use theory may seem a promising candidate. Another part of the appeal of a use theory comes from the difficulties of truth-referential theories. In particular it is proving hard to come up with a convincing theory of the ultimate referential links between language and the world.

So it would be nice if Horwich's use theory were promising. Drawing on the approach to semantics set out in *Coming to Our Senses* (1996: ch. 2), I shall argue that it is not. But these critical remarks do not start until part II. In part I I shall try, in what I hope is a sympathetic way, to develop Horwich's theory. I think that this development is necessary because Horwich does not give sufficient attention to the relation between language and thought.

## PART I

### 2. A Worthwhile Task?

There is an initial difficulty in doing semantics. What precisely is the task? There may seem to be a simple answer: The task is to say what meanings *are*, to explain their *natures*. It is thus analogous to such tasks as saying what genes, atoms, acids, echidnas, or pains are. However, we start the semantic task in rather worse shape than we do its analogues. With them, the subject matter of investigation is already identified relatively uncontroversially. This is because we already have a fairly good idea of what genes, atoms, etc., *do*, which is a basis for identifying them. Semantics does not start out like that. It is far from clear what counts as a meaning that needs explaining.<sup>2</sup> Thus Horwich takes the task to be to explain the property we attribute to a word in statements like the following:

'dog' means DOG.

<sup>1</sup> "The chief problem about semantics comes at the beginning. What is the theory of meaning a theory of?" (Higginbotham 1991: 271). "Meaning is notoriously vague" (Block 1986: 615).

<sup>2</sup> Although deflationist views of the truth term are clear enough, deflationist views of truth are not. My 2002 is an attempt to give a non-crude characterization of what the deflationist should say about truth. (My 2001 is a slightly expanded version.)

But what property of 'dog' does 'means DOG' identify? In particular, we would like to know what role the property we attribute is supposed to play. *What is it supposed to do?* For knowing what the alleged meaning property does is the key to identifying it.

We start semantics in the unusual position of having to specify a subject matter. We should not insist on great precision about this in advance of theory, but we do need some explication of our vague talk of "meanings." And we should surely insist on a subject matter *worthy* of investigation; we need an explication that is *not ad hoc*.<sup>3</sup>

In sum, the first step in semantics is to identify meanings. We can then move on to explain the meanings so identified.

### 3. Horwich's Theory

Horwich proposes a solution to the problem of identifying meanings: "the meaning-constituting property should be identified with whatever best explains the word's overall deployment" (p. 6).<sup>4</sup> By this deployment he has in mind the acceptance of sentences containing the word. So, for Horwich, what a meaning is supposed to do is explain that acceptance. Call this "the identification part" of his use theory. Having thus identified the properties that are the concern of semantics, the task is to explain their natures. What property of a word *does* best explain its overall deployment? Horwich's answer is: the property of being contained in "certain specified sentences" that are accepted (p. 6); so our acceptance of certain sentences containing the word explains our acceptance of any sentence containing the word. Call this "the explanation part" of his use theory.

Horwich does not really address the question of why his semantic task is worthwhile, although he may seem to in a couple of places.<sup>5</sup> Still I shall point out later that it clearly is worthwhile (sec. 5).

Horwich's presentation of his use theory, on which I have based the above summary, attends almost entirely to the linguistic. There is no explicit formulation of the use theory for the mental. Yet it is important to realize

<sup>3</sup> Bill Lycan has brought out the *ad hoc* problem wittily with his "Double Indexical Theory of Meaning":

MEANING =<sub>def</sub> Whatever aspect of linguistic activity happens to interest me now. (1984: 272)

<sup>4</sup> All such unidentified references are to Horwich 1998b.

<sup>5</sup> (a) At one point, Horwich says that his "account is not scientific explanation, but rather demystification" (p. 87). What he has in mind to demystify is the *theory* of meaning by freeing it from the "pseudo constraints" that he has earlier criticized (ch. 2). This is not addressing my question. We need to know why his semantic task is worthwhile even without the constraints. (b) At another point Horwich talks of demystifying meaning in the sense of explaining it (p. 58). Again we want to know why explaining meaning, in his sense, is worthwhile.

that Horwich intends his theory to apply as much to the mental as to the linguistic.<sup>6</sup> What then is his "mental use theory"? How does it identify and explain the meanings<sup>7</sup> of mental items?

Before answering, it helps to introduce some terms for those items. I shall talk of "mental sentences" and "mental words." In so doing I may seem to be embracing the rather controversial language-of-thought hypothesis. I do indeed believe that hypothesis but it does not play a role in this discussion. Take 'mental sentence' to be just a convenient way of referring to whatever it is in the mind that is the vehicle of the meaning we ascribe to beliefs and other thoughts. Similarly, take 'mental word' as a convenient way of referring to parts of that vehicle.

Using this terminology, we can formulate a mental use theory that is analogous to the linguistic one I summarized. We start with the identification part of the theory. The analogue of accepting a linguistic sentence is believing a mental one, "having it in the belief box." So we identify a mental word's meaning with that property of the word that best explains the presence of mental sentences containing the word in the belief box. That is what a meaning is supposed to do. Next we need to explain the nature of the meaning so identified, the explanation part of the mental use theory. What property of the mental word does best explain the presence of those sentences in the belief box? The analogue of the linguistic theory answers as follows: the meaning is the mental word's property of being contained in certain specified sentences in the belief box; the presence in the belief box of certain sentences containing the word explains the presence of any sentence containing it.

### 4. The Priority of the Mental

So, in effect, Horwich is proposing both a linguistic and a mental use theory. Does either have explanatory priority over the other? Not according to Horwich.<sup>8</sup> I shall argue that this is a mistake: he should give priority to the mental. I shall argue that his linguistic theory which seeks to identify and then explain the meanings of linguistic words depends on the prior identification and explanation of the meanings of mental words. The priority is a consequence of the linguistic theory's reliance on *accepting a sentence*.

Horwich identifies the meaning of a linguistic word with the property that best explains acceptance of sentences containing that word. What is it to

<sup>6</sup> See p. 44. The intention is not as clear in the book as it might be. I am confident that Horwich had the intention because he told me so.

At one point (p. 99), Horwich gives a brief nod to another theory that he sees as much the same as his use theory. Where his theory explains the nature of the property of 'dog' meaning DOG in terms of its use in sentences, the other theory explains the nature of the concept DOG—an abstract object—in terms of its appearance in propositions. Or "contentist": I shall talk of meanings.

<sup>8</sup> Again I am confident because he told me so.

accept a linguistic sentence? One answer is: to have in the belief box a mental sentence that the linguistic sentence expresses, a mental sentence "correlated with" the linguistic one. So, on this answer, the meaning of a linguistic word is identified as the property of the word that best explains the presence in the belief box of mental sentences containing the mental correlate of that word. But this is off track because what best explains the presence of a mental sentence, aside from the evidence, must be some property of *that mental sentence*. The linguistic word is beside the explanatory point. So this answer leads straight to the identification of word meanings at the mental level: a mental word's meaning is its property that best explains the presence of mental sentences containing the word in the belief box (as in my formulation above). The identification of the meaning of a linguistic word is then derivative: that meaning is the meaning of the mental word that the linguistic word expresses. And this priority of the mental in *identification* carries over to the *explanation* of the meaning so identified. We explain that meaning by discovering which property of the mental word *does* best explain the presence in the belief box of sentences containing it.

This argument for the priority of the mental starts from a very mentalistic account of accepting a sentence. We get the same priority, although not as swiftly, with the following less mentalistic account: to accept a linguistic sentence is to be disposed to assert the sentence in appropriate circumstances.

Why *would* a person in a linguistic community be disposed to assert a particular linguistic sentence, *T*, in appropriate circumstances? I propose the following explanation, "*E*":

- (I) Certain evidence causes her to believe a mental sentence with a certain meaning.
- (II) When she intends to express her belief, she is disposed to assert a linguistic sentence that has this meaning according to the conventions of her linguistic community.
- (III) *T* has that meaning according to the conventions of her linguistic community.

Against this background, consider Horwich's identification of a meaning with the property of a linguistic word that best explains acceptance of sentences containing the word. According to *E*, the meaning of a linguistic sentence is indeed its property that best explains its acceptance; see (III) particularly. Given that a word's meaning is the property it contributes to the meaning of a sentence containing it, we can infer that the word's meaning is indeed its property that best explains the acceptance of this sentence and any other containing it. So my explanation *E* confirms Horwich's identification of meanings. But, of course, the explanation also implies that this identifica-

tion does not go deep enough. *Why* does a property of a linguistic item that we have identified as its meaning, whether a property of a word or sentence, explain the acceptance of a sentence? *Only (a) because that property is the meaning of a mental item that is in the belief box; and (b) because the linguistic item conventionally expresses this mental meaning*. The acceptance story identifies a property as the meaning of a linguistic item only if we have already identified that property as the meaning of its correlated mental item. So according to *E* the identification of a linguistic meaning depends on the identification of a mental meaning. And, as a consequence, the explanation of the linguistic meaning will depend on the explanation of the mental. So this second account of accepting a sentence, like the first, results in the priority of the mental.

Consider the implications of this for the explanation part of Horwich's linguistic use theory. According to this theory a linguistic word's meaning is a certain "use regularity." The regularity in question is the word's property of being contained in certain linguistic sentences that are accepted. On either of our two accounts of accepting a sentence, the word will have this property only in virtue of expressing a mental word that has the property of being contained in the mental sentences in the belief box that the linguistic sentences express. In brief, assuming that the use property *is* the linguistic word's meaning, the word will have that meaning, only in virtue of its correlated mental word having it.

Perhaps there is a way of avoiding this conclusion by offering a behaviorist, hence entirely nonmentalistic, account of accepting sentences. But behaviorism is false (and is not embraced by Horwich). The conclusion seems inevitable on any nonbehaviorist account of acceptance. Consider, for example, a person's acceptance of 'Dogs often have fleas'. Clearly the explanation of this must see that acceptance as the result not only of the meanings of 'dog' and other words but *also* as the result of the person's assessment of the evidence for dogs often having fleas. How could these two factors be related together into the one explanation of acceptance if not via a mental meaning? The assessment of evidence is, after all, a mental matter.

If we are "intentional realists," hence nonbehaviorist, about the mind, then it is very hard to resist putting mind before language. This is not only the Gricean order of explanation, it is the folk order, reflected in the saying, "Language expresses thought."<sup>9</sup> I have not attempted to argue for this priority in general.<sup>10</sup> However, I have argued that Horwich's theory for the identifica-

<sup>9</sup> Horwich rightly criticizes (p. 52) the Gricean approach to semantics for offering no explanation of the meanings of thoughts.

<sup>10</sup> Because of the phenomenon of reference borrowing (deference), I think that the priority of the mental has to be qualified: although the mental has ultimate explanatory priority, it is not prior at all points in the explanation (Devitt and Stuebel 1999: 7.5).

tion and explanation of linguistic meanings must rest on the prior identification and explanation of mental meanings.

### 5. *The Identification of Meanings*

So I conclude that the identification of mental meanings is explanatorily prior. In this section I shall briefly assess the theory of this identification formulated on Horwich's behalf in section 3: a mental word's meaning is that property of the word that best explains the presence of mental sentences containing the word in the belief box.

This theory identifies the meaning of a mental word with a property that has a role in explaining why mental sentences containing it *are caused*. But doesn't this very same property also have a role in explaining what the sentences *cause*? Indeed, we have already noted that these sentences cause linguistic behaviors in virtue of their meanings. Thus, it is in virtue of a belief's meaning IT IS RAINING that an English speaker expresses that belief by asserting "It is raining." And why restrict ourselves to *linguistic* behaviors? Beliefs, and other thoughts, cause *nonlinguistic* behaviors in virtue of their meanings too: believing that it is raining not only leads me to say "It is raining" but also to take an umbrella. And causing nonlinguistic behaviors seems more fundamental (think of animals and babies, for example). So we should broaden the identification of meaning: a mental word's meaning is the property that plays a role in explaining not only why sentences containing the word are caused but also why they cause behaviors in general.<sup>11</sup> The meaning of a linguistic word is then identified as the meaning of the mental word that the linguistic word expresses.

In section 2 I claimed that we needed to identify meanings that are *worry* of study. Is the study of properties that we have just identified as meanings worthwhile? It surely is. Properties that play these key roles in the causation of thoughts and behaviors are playing scientifically important roles. And our account of them surely should be scientific, contrary to what Horwich suggests ("the use theory is not intended to be part of science"; p. 87).

This concludes my development of Horwich's use theory. I have presented a formulation of his mental use theory based on his explicit linguistic use theory. Horwich thinks that neither of these theories has explanatory priority. I have argued that he needs to give priority to the mental. Finally, I have suggested that the identification of meanings should be broadened to take account of their role as causes of behavior. So far as I can see, these devel-

<sup>11</sup> In my 1996 (2.5-2.6) I identified the meanings of mental sentences partly with their properties that cause behavior and partly with their properties that enable them to inform us about the world. I did not take account of the role of these properties in explaining the causes of mental sentences. I corrected this in 1997.

opments are compatible with Horwich's enterprise. This is not to say, of course, that he would accept them.

I shall now go on to criticize Horwich's linguistic use theory and my Horwich-based formulation of the mental use theory. I shall not criticize the identification parts of these theories; indeed, I think the broadened account is right. I shall be concerned with the explanation parts. I shall first consider Horwich's arguments for his theory; then, give some arguments against it.

## PART II

### 6. *Arguments for the Use Theory*

The explanation part of the mental use theory that I formulated by analogy with Horwich's linguistic theory is: the meaning of a mental word is its property of being contained in certain specified mental sentences in the belief box. This is its property that is supposed to best explain the presence in the belief box of any sentence containing the word and, I have just argued, the behaviors that those sentences cause. The explanation part of the linguistic use theory is derivative: the linguistic word's meaning is the meaning of the mental word that it expresses. So its meaning will be its property of being contained in certain specified linguistic sentences that are accepted, just as Horwich claims. These linguistic sentences will of course be those that express the mental sentences that determine the mental meaning.

Why should we believe these theories? Why should we suppose that a mental word's appearance in certain sentences best explains its causal role? Why not suppose, for example, that a word's referential properties best explain its role? Horwich begins his detailed presentation of the use theory with three particular use theories—for 'and', 'red', and 'true' (p. 45). These strike me as fairly plausible. But can plausible use theories be given for thousands of other words of very different sorts? In each case there is the problem of choosing *which* aspect of the causal role goes into the meaning. Precisely which mental sentences in the belief box are involved in the meaning of a mental word? Similarly, precisely which accepted linguistic sentences are involved in the meaning of a linguistic word? These questions raise the notorious problem of "the principled basis" for choosing meaning-constituting properties. Horwich's idea for solving this problem is, of course, that the meaning-constituting aspect of a word's causal role is the one that *best explains* the other aspects. But in the absence of a wide range of plausible examples of such explanations we need to be persuaded that *there is* some particular aspect of each word's causal role that does this explanatory job. I don't think that Horwich says nearly enough to persuade us of this.

Before considering what Horwich says for the use theory, it will help to sketch an alternative theory for comparison. The alternative is a "moderate"

truth referentialism along the following lines.<sup>12</sup> "Primitives" get their meanings from referential relations explained by some sort of direct causal link to reality, an informational, teleological, or historical-causal link, or some combination of these; proper names and natural kind words are likely primitives. Other words get their meanings from referential relations explained by their "definitional" links; the words are inferentially associated with others that determine their reference; they are covered by "description" theories of reference; 'bachelor' is a likely example.

In assessing the relative merits of the use theory, it is important to note that, for many words, the theory will be hard to distinguish from moderate truth referentialism. The words in question are ones like 'bachelor' that truth referentialism treats as nonprimitive and covered by a description theory. Thus, where truth referentialism explains the meaning of 'bachelor' by its reference-determining association with 'unmarried man', the use theory explains its meaning by its appearance in 'A bachelor is an unmarried man'. For such words, the only significant difference between the two theories is in their attitudes to reference: truth referentialism is committed to a substantial reference relation, the use theory, to deflationary reference. According to the use theory, meaning trivially determines reference; according to truth referentialism, meaning strongly determines reference (see pp. 68-71 on this distinction.) In light of this, when we attend to words like 'bachelor' we are unlikely to find decisive evidence favoring one theory over the other (independent, of course, of evidence for or against a deflationary theory of reference).

To find decisive evidence we should look to the handling of truth referentialism's primitives. Truth referentialism tries to explain these words not by their associations with other words but by their direct causal links to reality. One choice for a use theory is to do likewise, thus also treating them as primitives.<sup>13</sup> But then the theory faces two challenges. First, it needs to distinguish itself from truth referentialism. Truth referentialism holds that these meaning-constituting causal links explain reference. The use theory needs to say why they don't (because the use theory holds that reference is deflationary and not open to this sort of explanation). Second, truth referentialism has had difficulty finding causal links that can plausibly be seen as

<sup>12</sup> I urge a view of this sort in 1996 and Devitt and Sterelny 1999. The naturalistically-inclined truth referentialist has to choose between some such moderate view and the extremes of semantic atomism (every word is a primitive).

<sup>13</sup> I take it that this is what Horwich is doing in the following use theory:

the explanatorily fundamental acceptance property underlying our use of 'red' is (roughly) the disposition to apply 'red' to an observed surface when and only when it is clearly red. (p. 45)

This explains the meaning of 'red' (largely, if not entirely) in terms of a certain causal relation between speakers and red surfaces.

meaning-constituting. Indeed, a large part of the appeal of the use theory comes from this problem for its rival (sec. 1). The use theory faces the challenge of doing better: showing that its non-reference-determining causal links better explain meaning than truth referentialism's reference-determining ones. The other choice for a use theory—a much less appealing one in my view—is to insist that these words are *not* primitives, that they are to be explained by their association with other words (in effect). But then well-known truth-referentialist arguments to show that the words *are* primitives, and are not to be handled by description theories, are likely to count against the use theory. These matters will loom large in the next section.

Horwich offers seven arguments for his use theory, two considerations favoring the theory over alternatives, and answers to twenty four objections. I think that Horwich greatly overestimates the success of the use theory and the difficulties of its rivals, particularly a moderate truth referentialism. Hence he gives a misleading picture of the "state of play." Furthermore, Horwich's discussion does not provide significant support for his solution to the principled basis problem.

In this section I shall assess the seven arguments for and the two favoring considerations.

"(1) The Univocality-of-'Meaning' Argument": Where Grice distinguished a "natural" sense of 'mean' exemplified in "Black clouds mean rain" from a "nonnatural" sense exemplified in "'Dog' means DOG" Horwich insists on the univocality of meaning: the notion "is, roughly speaking, the notion of *indication*" (p. 47). A person's use of the word 'dog' indicates the presence in his mental state of the concept DOG. But this does not favor the linguistic use theory over *any theory that gives priority to the mental*: on any such theory a word expresses a concept and hence can be taken to indicate the concept's presence. The theoretical action is then over the concepts (mental meanings). In this action, univocality gives no preference to the use theory of concepts over any other.

"(2) The Explanation Argument"; the "explanatory power" consideration:

The explanation argument is:

it is indeed quite clear (as we have just seen) how the total use of a word might be derived...from a basic "law" of use—whereas it is relatively unclear how any other sort of "property" would [do]. (p. 47)

Similarly confident claims have been made earlier:

use theory...alone...is able...to account for the fact that we apply each of our words as we do in part because of what we mean by it. (p. 6)

[the use theory] accommodates the Use Constraint just about as well as anything could. (p. 40); the Use Constraint is "to account for the relationship between the meaning of a word and its use.")

And the explanatory power consideration (p. 53), alleged to favor the use theory over alternatives, is exemplified primarily in accounting for the overall use of words. The puzzling thing is that, on this matter that is so central to the explanatory part of Horwich's use theory, there is just one brief argument. The argument concerns 'bachelor' (pp. 53-54). Yet, for reasons just indicated, the argument over this word does not favor the use theory over moderate truth referentialism. When we look for support for Horwich's solution to the principled basis problem we find almost nothing. It looks as if Horwich thinks that these claims about explanation are *intuitively obvious*. So it is supposed to be obvious that our acceptance of certain sentences containing a word always explains our acceptance of any sentence containing it, and that other things don't. But why should the causal structure of the world be any more obvious here than it is almost anywhere else?

Far from being obvious, I shall soon argue that these claims are false (although certainly not *obviously* false!).

"(3) **The Meaning-Attribution Argument**": "we do establish what is meant by a word by observation of how it is used" (p. 48). And so we do. But this reflects only the relatively uncontroversial fact that the meaning of a linguistic word is the *result* of its use: the word has that meaning by convention, a convention arising from its regular use in a linguistic community to express that meaning. Thus, Italians use 'cane' to express DOG thoughts because of a convention of so doing. This does not support any theory of the meaning of 'cane' because it does not throw any light on the *nature* of the resulting meaning. So it does not support the view that the meaning which is the *result* of the use of 'cane' is to be identified with that use. More particularly, it does not show that the *acceptance* properties of 'cane' have anything to do with its meaning. Furthermore, it shows nothing about mental meaning: it is compatible with absolutely any theory of what makes DOG thoughts DOG thoughts. So far as this argument is concerned, it could be the case that the meaning of 'cane' and the mental word it expresses is the property of referring to dogs. If this were so, what we would be looking for, in observing how the word is used to establish its meaning, would be evidence of reference.

"(4) **The Synonymy Argument**": We need to explain why speakers of a language will usually be prepared to substitute synonym for synonym in any sentence they accept. The use theory can do this easily, Horwich claims, but a referential theory cannot (p. 49). But it is not clear why any theory, except perhaps an extreme referential theory, could not explain this. Thus a description theory of 'bachelor' is as well placed as a use theory to explain

our readiness to substitute 'unmarried man' for 'bachelor'. This is an example of a phenomenon that a description theory, committed to a substantial referential relation, can explain as readily as a use theory, committed only to deflationary reference.

"(5) **The Implicit Definition Argument**": "One may introduce a new term, 'f', and give it a meaning simply by accepting a body of postulates, '#f', containing the term." What does meaning have to be like for this to be so? Horwich claims that "it is hard to think of a plausible alternative" to the use theory (p. 50). In fact, it is very easy to think of one: a description theory according to which the postulates fix inferential associations that (strongly) determine the reference of 'f'. Once again, a description theory can do as well as a use theory.

"(6) **The Translation Argument**": "The way in which we operate with manuals of translation... is explained—and can only be explained—by means of the use theory of meaning." For, a manual "is used as a device of 'expectation replacement'... instead of asserting *our* sentences, we assert the translations of them, supposing that this will generate the same relevant expectations in the audience as our sentences would at home" (p. 50). But why is "expectation replacement" the purpose of the manual? By ascribing meanings to any person's words we gain knowledge of her thoughts which we can then use to explain her behavior (and give us information about the world). Now if the person speaks our language the ascription is easy. If not, we need a translation manual. So the manual's function is to match a foreigner's words with ours so that we can figure out what she's thinking when she speaks and she can figure out what we are thinking when we speak; expectation replacement is derivative from this function. And any theory of meaning can account for this function.

"(7) **The Pragmatic Argument**": "in so far as [the use theory] explains why she should seek manuals of translation, it explains, *a fortiori*, why it is valuable to possess the concept of translation and therefore the concept of meaning; hence it accounts for us having those concepts" (p. 51). But if my story of the function of the manual is right, any theory of meaning can account for our having those concepts.

**The "generality" consideration**: The other consideration alleged to favor the use theory over alternatives is "its generality." Thus, a "definition theory" cannot be the whole story because "there would have to be a certain residue of indefinable primitives." "And the informational and teleological theories cannot aspire to deal with the terms of logic and mathematics" (p. 53). But is it really clear that the use theory is more general than moderate truth referentialism? And if it were, would that be much of an advantage? (But I do think that the use theory has an advantage over its rivals in the handling of the terms of logic and mathematics.)

In sum, I think that the seven arguments and two considerations greatly underestimate the possibility that theories other than Horwich's linguistic use theory, or the mental use theory based on his, could explain the phenomena in question. And they offer very little to support Horwich's solution to the principled basis problem: that for each word its involvement in some particular set of sentences explains its causal role. So the arguments do not mount a persuasive case for his use theory.

#### 7. *Arguments Against the Use Theory*

Next, consider objection 13, **Communal Meanings**, and objection 15, **Names**, arising from the work of Saul Kripke, Hilary Putnam, and Tyler Burge: "members of a linguistic community typically mean exactly the same as one another by a given word, even when their uses of it diverge" (pp. 85-86). This reflects an argument that referentialists have made against description theories: competent users of, say, a name, are simply too ignorant or too wrong to give an accurate description of its bearer; a description theory places too great an epistemic burden on speakers. The argument is, in effect, that names are "primitives."

I anticipated earlier that the struggle between truth-referentialist and use theories would come down to the handling of truth referentialism's primitives. And so it proves.

Horwich responds to these objections by allowing a person to mean something by a word even though she does not use it "precisely in accordance with the regularity that fixes the meaning of the word." How is this possible? Horwich borrows from Putnam:

What is needed is, first, that there are acknowledged experts in the deployment of the term—experts whose usage is determined by some such regularity; second, that the individual is disposed to defer to the experts—i.e. to accept correction by them; and consequently, third, that his use of the term conforms to that regularity at least to some extent. (p. 86)

The most important thing to note about this move is that it requires an *awful lot* more explanation and defense. The move is similar to attempts to save the description theory of names. So we can draw on responses to those attempts.<sup>14</sup>

Consider proper names. To meet Horwich's second requirement, the user of a name has to acknowledge her ignorance and hence be prepared to defer. But surely many ignorant users of a name do not acknowledge their ignorance. On Horwich's theory these users will not be using the name with the same meaning as the deferrers and the experts. My guess is that most users of most names will be in that category. Next, the requirement demands that each deferrer identify experts to defer to. How? They cannot be identified simply as

experts on the meaning of the name, on pain of circularity. It is surely unlikely that most deferrers will be able to manage the required identification. The smart modern way to do so with a famous name is to consult an encyclopedia. But what about the time before there were any encyclopedias, or the millions who have never heard of encyclopedias? And what about the millions of nonfamous names? In brief, even if there are acknowledged experts, many users of many names will have no way of identifying them. Finally, Horwich's first requirement is that there be acknowledged experts. But surely the ignorant will often defer to someone that they think is an expert who in fact is not. In sum, people will often not defer where they should; they will often try to defer but fail; they will often defer to a non-expert. Horwich's use theory places far too heavy an epistemic burden on speakers. The very same considerations that were devastating for description theories of names are likely to be devastating for Horwich's use theory.

The problem is most obvious with names but arises with any word that might plausibly be seen as primitive and not open to a description theory in a truth-referentialist semantics; so it arises for natural kind words, perhaps for artificial kind words like 'sloop' and 'sofa', medical words like 'arthritis', socio-legal words like 'contract'. These are difficult matters on which there is plenty of room for argument. Until he has joined the argument, Horwich's theory is, at best, seriously incomplete. At worst, it will founder because of ignorance and error problems like those I have just briefly indicated for his theory of names.

In the face of these problems it is tempting for the use theorist to go in for more theory borrowing, this time from those who have attempted to explain primitives in terms of an informational, teleological, or historical-causal link to reality, or some combination of these. But then the use theorist would face the two challenges mentioned in the last section.

The first challenge is to distinguish such a use theory from truth referentialism. If these causal links borrowed from other theories really do constitute meaning why do they not also determine reference as those other theories claim? If they do determine reference then, of course, this use theory has collapsed into a truth-referentialist one.

The second challenge is to do a better job than truth referentialism in locating the meaning-constituting causal links. It is not sufficient for a truth referentialist to talk vaguely about direct causal links to reality constituting a meaning: the devil is in the details. The devil is there also for a use theory. A use theory that appeals to causal links to reality in explaining a word's meaning must say precisely which causal links. We not only need assurance that this use theory is not collapsing into a theory of reference, we need to be convinced that it really can explain meaning with these causal links. So far as I know, use theorists have not attempted this task. Whereas truth-referential-

<sup>14</sup> See, e.g., Devitt and Sterelny 1999: 3-4.

ists have attended a lot to the details, use theorists have yet to do so. So it remains to be seen whether the difficulties for the use theory will be any less than those for truth referentialism.

Finally, Horwich's mental use theory is, in effect, a functional-role theory. Elsewhere, I have argued against functional-role theories in general (1996: particularly, 4.9-4.10). I summarize. In our ordinary ascriptions of meanings to thoughts the only properties we ever ascribe to a mental word are referential ones, either a property of referring to something or a property that determines reference; those and only those properties are common and peculiar to the parts of thoughts to which we ascribe a meaning using clauses like 'that...wheel...', 'that...Cicero...', and 'that...atom...'. No functional-role property, *except one that determines reference* (hence covered by a description theory), is ever ascribed. If this argument is right, the properties that folk and social scientists ascribe to thoughts to explain behavior (and guide us to the world) are referential. And these explanations are mostly successful (*pace* Churchlands). So the argument provides powerful evidence against what Horwich seems to find obvious: that our acceptance of certain linguistic sentences containing a word explains our acceptance of any sentence containing it. Similarly, it provides powerful evidence against the view that our believing certain mental sentences containing a word explains our believing any sentence containing it. What explains this acceptance and belief, and our behavior in general, are the referential properties of words. So those properties are meanings.

This referentialism goes against the popular internalist intuition that properties partly constituted by matters external to the mind *could not* explain behavior. But then Horwich's use theory goes against the intuition too. In any case, the intuition is just that: an intuition. If my argument is right, what we are ascribing, day in and day out, in largely successful attempts to explain behavior, are referential properties. This is evidence that these properties do explain behavior and so are meanings. It is evidence that the internalist intuition is wrong. Indeed, if we generalize the internalist intuition to science as a whole it seems to me clear that it is wrong.<sup>15</sup>

In part I, I proposed some developments of Horwich's use theory that give priority to thoughts. In this part I have argued, first, that Horwich greatly overstates the case for his linguistic use theory and, hence, for the mental use theory based on that theory. I have argued, second, that the arguments from ignorance and error against description theories of reference can be adapted to provide a strong case against these use theories. I have argued, third, that a tempting development of the use theory that appeals to direct causal links to reality would risk both the collapse of the theory into truth referentialism and

the difficulties that have plagued truth referentialism. The most important point to emerge from these three arguments is that we need a lot more details of the use theory before it can be taken as a serious rival to truth referentialism. Finally, a consideration of our ordinary thought ascriptions provides evidence against any functional-role theory, including a use theory.

So I do not think that we should be persuaded by Horwich's use theory. Still, his book is a valuable contribution to the ongoing debate on the difficult matter of the nature of meaning.<sup>16</sup>

#### REFERENCES

- Block, Ned 1986. "Advertisement for a Semantics for Psychology." In French, Uehling, and Wettstein 1986: 615-78.
- Brandom, Robert 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge MA: Harvard University Press.
- Devitt, Michael 1996. *Coming to Our Senses: A Naturalistic Program for Semantic Localism*. New York: Cambridge University Press.
- Devitt, Michael 1997. "On Determining Reference." In *Sprache und Denken / Language and Thought*, Alex Burri, ed. New York: Walter de Gruyter: 112-21.
- Devitt, Michael 2001. "The Metaphysics of Truth." In *The Nature of Truth*, ed. Michael Lynch. Cambridge, MA: MIT Press: 579-611.
- Devitt, Michael 2002. "The Metaphysics Deflationary of Truth." In *What is Truth?*, ed. Richard Schantz. Berlin & New York: Walter de Gruyter: 60-78.
- Devitt, Michael, and Kim Sterelny 1999. *Language and Reality: An Introduction to the Philosophy of Language*, 2nd edn. (1st edn. 1987). Oxford: Blackwell Publishers.
- French, Peter A., Theodore E. Uehling, Jr., and Howard K. Wettstein, eds. 1986. *Midwest Studies in Philosophy Volume X: Studies in the Philosophy of Mind*. Minneapolis: University of Minnesota Press.
- Higginbotham, James 1991. "Truth and Understanding." *Iyyun, the Jerusalem Quarterly* 40: 271-88.
- Horwich, Paul 1998a. *Truth*, 2nd edn. (1st edn. 1990). Oxford: Clarendon Press.
- Horwich, Paul 1998b. *Meaning*. Oxford: Clarendon Press.
- Lycan, William G. 1984. *Logical Form in Natural Language*. Cambridge, MA: MIT Press.

<sup>16</sup> A draft of this was delivered in a course, "Meaning," given jointly by Paul Horwich and me at the Graduate Center in Fall 2000. I am indebted to Horwich's detailed responses, and the comments of the class, for many changes.