

MICHAEL MCKINSEY

## APRIORISM IN THE PHILOSOPHY OF LANGUAGE

(Received 9 September, 1986)

In this paper, I will try to motivate, clarify, and defend a principle in the philosophy of language that I will call "apriorism." Roughly stated, apriorism is the principle that the meanings of words are knowable a priori. Stated this way, the principle requires considerable clarification, and I will try to provide a more exact version below. But even from this rough statement, I think it is fairly clear that some principle of this sort is taken for granted by most philosophers of language and by many linguists. Certainly, philosophers in particular seem to proceed with their conceptual investigations as if apriorism were true. For the evidence that they rely on to confirm or disconfirm their hypotheses concerning the meanings of particular sentences, words, and classes of words, seems to consist almost solely of evidence that is provided by their own linguistic intuitions and dispositions.

But even though apriorism seems to be commonly presupposed, some important recent work in the theory of reference seems to support conclusions that entail its denial. In particular, Hilary Putnam's famous "Twin Earth" example can be used to construct a convincing argument against apriorism. (Putnam, 1975.) Putnam originally used this example to argue against the Fregean doctrine that to know the meaning of a word is to be in a state that is psychological "in the narrow sense," where a state is psychological in the narrow sense only if its existence entails the existence of no contingent object besides the person who is in the state.<sup>1</sup>

This Fregean doctrine resembles apriorism, but the two principles are not the same. Thus unlike the Fregean doctrine, apriorism allows that a person's knowing a certain word's meaning might entail the existence of a contingent object other than the person, provided that the object in question is one whose existence is knowable a priori. If I

am right, this apparently small difference in the two principles proves to be quite significant, since it will enable me to show that, while Putnam's Twin Earth argument may refute the Fregean doctrine, it does not refute apriorism.

Though my chief aim is to defend apriorism against the Twin Earth argument, my defense unfortunately requires a large amount of preliminary stage-setting, for which I ask the reader's indulgence. I will begin by trying to motivate and clarify the version of apriorism that I take to be true and that I wish to defend.

#### PART I. APRIORISM

##### 1. *When Is Knowledge A Priori?*

Although I think that many, and perhaps most, philosophers would be inclined to accept the description 'a priori' as applying to knowledge of meanings, it is difficult to say precisely how the description should be understood in this context. It is common to take logic and mathematics as paradigms and characterize a priori knowledge as in some sense purely *conceptual* in nature. But for reasons that will become clear later, knowledge of meanings cannot be understood this way. As we shall see, knowledge of meanings instead resembles knowledge of one's own existence or of one's own mental states, and of course knowledge of this latter sort is not conceptual in nature, since it is knowledge of contingent facts.

Moreover, unlike knowledge in logic and mathematics, knowledge of meanings is obtained only through the liberal use of *induction* as well as deduction. For typically, an hypothesis concerning the meaning of an expression or class of expressions is an inference to the best explanation, where the facts to be explained are provided by linguistic intuitions.

We tend to classify knowledge as a priori if it is knowledge that we can obtain "just by thinking," or in other words, knowledge that is not based upon perceptual observation or empirical investigation. It is because knowledge of meanings is like this that we are inclined to say that it is "a priori." Although I cannot give an adequate account here of what it means to obtain knowledge "just by thinking," per-

haps the following characterization will suffice for my present purposes.

I will say that a person's knowledge is a priori if and only if it would remain knowledge even if the person were radically deceived in his assumptions and inferences concerning the existence and nature of the physical world that is external to his mind. Thus on this characterization if a piece of my knowledge is a priori, then I would still have this knowledge even if I were a brain in a vat and had been systematically deceived by a mad scientist; I would still have this knowledge even if I were a nonphysical mind in a nonphysical world and had been systematically deceived by Descartes' evil genius.

On this account, there can be a priori knowledge of both necessary and contingent truths. For on this account, or so it seems to me, I know a priori that I exist, that the present time exists, and that various of my present mental experiences and states exist. These are of course things that might not have existed, but it seems to me that I have a priori knowledge of their existence, for these things would still exist and I would still know that they do, even if I were radically deceived in my assumptions about the external world. On the other hand, my characterization of a priori knowledge does not imply that all such knowledge is incorrigible, or known with some special kind of certainty. No doubt some such knowledge is in fact incorrigible in the sense that one's believing it entails its truth. (For example, my believing that I exist entails that I do.) But much knowledge that is a priori on my account is certainly not incorrigible, and is fallible in just the way that other types of inductively based knowledge are fallible.

## 2. *Knowability*

Perhaps we now have a slightly clearer idea of the sense of 'a priori' in which it may be true to say that meanings are knowable a priori. Still, there is a further serious problem that stands in the way of understanding this principle. When we say that something is knowable a priori, we mean that it can be known a priori, or that it is *possible* that it be known a priori. But what sense of 'possible' do we have in mind here?<sup>2</sup> It seems unlikely that we mean mere logical

possibility. For it seems to be at least logically possible that there should be an omniscient being – God perhaps – who knows *every* true proposition a priori, including ordinary empirical propositions about the physical world. If this is so, then in order for it to be of any interest that something is knowable a priori, this cannot mean merely that there is some logically possible situation in which it is known a priori by some being or other.

For the sake of definiteness, when I say that something is knowable a priori I will mean that it is *humanly* possible to know it a priori. I can limit the scope of the subject in this way, since my aim is to defend the principle that meanings are knowable a priori by the speakers of languages in which words have these meanings, and I am really only concerned with humanly speakable languages. Thus I will confine my attention to such languages and assume that if a meaning of a word is knowable a priori then there is some world that is possible relative to the laws of human psychology in which some human being knows that meaning a priori.

If we hold that it is humanly possible to know meanings a priori, then we endorse a stronger thesis than if we hold merely that it is logically possible to know meanings a priori. But a proponent of apriorism should want to endorse a still stronger thesis. Let us call a world that is possible relative to the laws of human psychology, a “humanly possible world.” So far, we have apriorism committed only to

- (1) Necessarily, if a word has a given meaning in a humanly speakable language, then there is some humanly possible world in which some human being knows a priori that the word has that meaning in that language.

Now (1) is a fairly weak principle. For it is compatible with (1)’s truth that no speaker of any given actual language ever has had or will have the *capacity* or ability to know a priori the meaning of a single word of that language. All that (1) requires is that it be psychologically possible for humans to *acquire* such a capacity, and to then go on and exercise it.

But why would a proponent of apriorism think that (1) is true? Surely, one would think that it is psychologically possible to acquire

a priori knowledge of the meanings of words only because one also thought that speakers of natural languages often do in fact have the capacity to acquire such knowledge. Thus a proponent of apriorism would endorse (1) only because (1) follows from the stronger thesis

- (2) Necessarily, if a word has a given meaning in a language that has human speakers, then at least some of those speakers have the capacity to know a priori that the word has that meaning in that language.

The motivation behind the apriorist's endorsement of (2) comes in turn from a certain view of what it is to understand the words of a language. On this view, a speaker has mastery of a word's meaning and can use it with understanding only when the speaker has in some sense "internalized" that meaning, or perhaps, has "internalized" the semantic rule of his language by virtue of which the word has that meaning. But it is precisely by internalizing a word's meaning that a speaker acquires the capacity to know a priori what that meaning is. For this process of internalizing a word's meaning is one and the same as the process whereby a speaker acquires certain linguistic propensities; and it is knowledge of these propensities that forms the basis of the speaker's a priori knowledge of the word's meaning.

### 3. *Apriorism Qualified*

It is tempting at this point to simply identify apriorism with the principle (2). However, though (2) comes close to capturing what the proponent of apriorism has in mind, it is not quite correct. To see why it may help to first see what is wrong with the similar but slightly stronger thesis:

- (3) Necessarily, if a word has a given meaning in a human language, then any speaker of that language whose uses of the word have that meaning has the capacity to know a priori that the word has that meaning in that language.

(3) is false for the simple reason that a given word, and hence a speaker's use of the word, may have a certain meaning in a language, even though the speaker misunderstands what that meaning is. In this

sort of situation, a speaker's *words* can mean something other than what the *speaker* means by the words.<sup>3</sup> For instance, suppose that Professor Smith, having confused the meaning of the word 'philanderer' with that of the word 'philatelist' says to the visiting dignitary, "The Dean is a great philanderer, you know." What Professor Smith *means*, of course, is that the Dean is a stamp-collecting enthusiast. But unfortunately what the Professor has *said*, what his words mean, is that the Dean is a profligate womanizer. Moreover, contrary to what (3) implies, the Professor does not have the capacity to know a priori that this is what his words mean.

A token or use of a word can only mean semantically what the word itself means in the language being spoken. Thus Professor Smith's use of the word 'philanderer' can only mean what 'philanderer' means in English, namely, 'profligate womanizer'. But the Professor cannot discover a priori that this is what his use of 'philanderer' means. For by consulting his own linguistic dispositions, he can come only to the mistaken conclusion that a philanderer is a person who collects stamps. To find out what the word 'philanderer' actually means in English, Professor Smith must consult a dictionary, question his fellow speakers or use some other method that does not yield a priori knowledge. Thus, a speaker's use of a word can have a given meaning, even though the speaker does not have the capacity to know a priori that the word has that meaning. So (3) is false.

One who admits that (3) is false for the reason just given might still insist that (2) is true nonetheless. For (2) just says that if a word has a meaning in a language, then at least *some* speakers of the language can know that meaning a priori. And it is plausible to suppose that this is true in spite of (3)'s falsity. After all, a speaker who is ignorant of a word's meaning can use the word with that meaning only because there are *other* speakers of the same language who are *not* ignorant of the word's meaning. For surely, if there were not at least *some* speakers of the language who correctly understood the word's meaning, then the word simply would not *have* that meaning in that language. But then these other speakers must have internalized the correct meaning and so, one might think, these speakers must have the capacity to know that meaning a priori.

But the last step of this argument is mistaken. It is true, I think, that a word could not have a given meaning in the language of a group of speakers unless at least some of those speakers correctly understood the word as having that meaning. But it does not follow that those speakers who correctly understand the meaning of the word would have the ability to know a priori that the word has that meaning. Such a speaker would have internalized the word's meaning all right, and as a result, the speaker would have acquired linguistic propensities, a priori knowledge of which would lead the speaker to the correct conclusion as to what the word means. But the speaker could not know a priori that this conclusion is correct. For the most that the speaker can know a priori concerning the word's meaning is that he has internalized that meaning. He cannot know a priori that the meaning he has internalized is the meaning that the word in fact has in his language. To reach this last conclusion, the speaker must assume that his understanding of the word's meaning is *correct*. The speaker must assume, for instance, that he is not in the same sort of situation as Professor Smith, who upon examining his own linguistic propensities regarding the word 'philanderer' reached a false conclusion about this word's meaning. But the speaker cannot know a priori that he is not in this sort of situation even if he really is not.

The most that a speaker can know a priori concerning a word's meaning is what he himself means by the word. By considering one's own linguistic intuitions and propensities, one can sometimes succeed in reaching an adequate generalization on a priori grounds concerning what the word *would* mean if it meant what one means by it. Speakers can, in other words, find out a priori what meaning they have internalized, independently of whether the word in question actually has that meaning. This sort of a priori knowledge is possible because it is a species of a priori knowledge concerning one's own mental states. It is like the a priori knowledge that people have of their own desires, intentions and beliefs.<sup>4</sup>

We have seen that a word has a given meaning in the language of a group of speakers only if at least some of those speakers correctly understand the word's meaning, only if, in other words, that meaning is what some of those speakers themselves mean by the word. We have also seen that people do in general have the capacity to know a

priori what they mean by a word. Thus we have

- (4) Necessarily, if a word has a given meaning in a language that has human speakers, then at least some of those speakers have the capacity to know a priori that that meaning is what they themselves mean by the word.

I suggest that (4) captures the sense in which it is true to say that meanings are knowable a priori, and so (4) expresses the version of apriorism that I wish to defend below. In keeping with the motivation behind (4), it will be assumed from now on that a given word's meaning is "knowable a priori" if and only if it is humanly possible to know a priori that that is what one means by the word.

Though (4) is weaker than other possible expressions of apriorism such as (2) and (3), it is still a thesis that is far from trivial. For (4) still places a significant constraint on the meanings that the words of human languages can have. Nor have we managed to avoid Putnam's Twin Earth argument merely by opting for a version of apriorism that is weaker than any version to which the argument applies. As we shall see, Putnam's example poses the same problem for (4) as for the other versions of apriorism.

Moreover, (4) is strong enough to provide a justification for the standard methodology in contemporary philosophy of language. Thus, when a philosopher investigates from his armchair the meaning of a given expression, by considering in some detail his own linguistic intuitions and propensities concerning examples involving the expression in question, he is in the first instance attempting to construct an a priori theory of what he means by the expression, or as we sometimes say, a theory of what the expression means "in his idiolect." As anyone who has tried to do it can attest, this sort of task is far from being easy or trivial, and can sometimes involve a good deal of detailed and complex clarification and theory construction. So the fact that the subject matter may be idiosyncratic does not make the investigation trivial.

Nor does the possible idiosyncrasy of the subject matter make the investigation irrelevant to the main goal, namely, that of learning something about a given expression's meaning in a public language. For once one has obtained a priori knowledge of what one means by



an expression, this knowledge, together with the assumption that one's understanding of the expression is correct, provides a conclusion as to what the expression means in the public language one speaks. The assumption that one's understanding is correct, though not knowable a priori, is of course very often justified and known to be true. And when the assumption is known to be true, one's a priori knowledge of what one means by an expression in turn provides knowledge of what the expression itself means.

## PART II. TWIN EARTH AND NATURAL KINDS

### 4. *Initial Statement of the Twin Earth Problem*

In Putnam's familiar example, we are to imagine that somewhere in the galaxy there is another planet that is nearly an exact duplicate of Earth. This other planet, which Putnam calls 'Twin Earth', is exactly similar to Earth in its history, appearance and physical structure. There is even, for every person and object on Earth, an exact duplicate, or *doppelgänger*, on Twin Earth. The *doppelgängers* of those who speak English on Earth also speak English on Twin Earth, and the histories and mental lives of these *doppelgängers* are qualitatively indistinguishable from those of their counterparts on Earth. The major difference between Earth and Twin Earth is that the liquid found in the lakes and rivers of Twin Earth, and called 'water' by the English speakers of Twin Earth, has an entirely different chemical structure from the liquid that we on Earth call 'water'. While the latter stuff is composed of  $H_2O$  molecules, the stuff called 'water' on Twin Earth is instead composed of, say, XYZ molecules. Otherwise, though, the two liquids are indistinguishable: both are colorless, odorless, thirst-quenching, found in rivers and lakes, fall from the sky as rain, etc..

Now it seems obviously correct to say, as Putnam does, that what the Twin Earthians call 'water' is not water. For water is  $H_2O$ , and what the Twin Earthians call 'water' is not  $H_2O$  but XYZ. Thus the word 'water' when used by us on Earth has a different extension than it does when used by our *doppelgängers* on Twin Earth. But then, given that a general term's extension is determined by its meaning,

there must also be some difference between the meaning that ‘water’ has on Earth and the meaning it has on Twin Earth. The problem is to account for how this difference in meaning is possible, given the extreme qualitative similarity of Earth and Twin Earth.

The difference clearly cannot be accounted for if we assume that the meaning of ‘water’ as we use it on Earth specifies its extension in a purely qualitative way. For instance, we cannot suppose that ‘water’ has the following sort of definition:

- (5)  $x$  is water  $\equiv_{df}$   $x$  is colorless, odorless, thirst-quenching liquid of the kind that fills up lakes, rivers, and oceans, and that falls from the sky as rain.

The problem is that this specification is satisfied by both samples of  $H_2O$  on Earth and samples of XYZ on Twin Earth. So if ‘water’ as we use it had this meaning, we would get the false consequence that XYZ is water after all.

Nor can the meaning of ‘water’ as we use it be given via a specification of water’s chemical structure. For although water is  $H_2O$ , ‘water’ does not *mean* ‘ $H_2O$ ’. ‘Water is  $H_2O$ ’ is not an analytic truth, and ignorance of the fact that water is  $H_2O$  does not betray one’s ignorance of the meaning of the word ‘water’.

It seems that in order to specify water so as to distinguish it from what Twin Earthians call ‘water’, a person with no scientific knowledge would have to mention some *object* to which water, but not XYZ, bears a certain relation. For it seems that the only differences that would exist between water and XYZ, besides the chemical differences, lie in the distinct objects to which the two kinds of liquids would be related. Thus water, but not XYZ, is found in the lakes and rivers of *Earth*; water, but not XYZ, is stuff that *we* (the inhabitants of Earth) have experienced; and so on.

Perhaps, then, ‘water’ could be defined by mentioning some particular object to which water, but not XYZ, bears a certain relation. For instance, we might try a definition like:

- (6)  $x$  is water  $\equiv_{df}$   $x$  is colorless, odorless, thirst-quenching liquid of the kind that is found on Earth.

I assume that in (6) the name ‘Earth’ is used as a genuine term that

refers directly to, without describing, the planet Earth. By a “genuine term,” I mean a term whose sole semantic function is that of introducing a referent into what is said by sentences containing the term. Given this way of understanding (6), this definition seems to avoid the problem raised by the Twin Earth example, since in this example, XYZ is not a liquid that is found on Earth.

Let us say that an “objectual meaning” is a meaning that can only be expressed by use of a genuine term that refers to a particular concrete object. Such a meaning essentially involves a particular object, or has that object “as a constituent.” For instance, if we identify the meaning that (6) ascribes to ‘water’ with the relational property expressed by (6)’s *definiens*, then this property would be an example of an objectual meaning that has the planet Earth as a constituent.

What the Twin Earth case seems to show, then, is that there are objectual meanings. And it is here that the example raises such a serious problem for apriorism, for objectual meanings are not in general knowable a priori. Consider, for instance, the relational property expressed by (6)’s *definiens*. This property could not exist unless the planet Earth did, and so one could not know that this property exists without knowing that Earth exists. But then, since of course one cannot know a priori that Earth exists, one cannot know a priori that the property in question exists. Hence one could not know a priori that this property is the meaning of the word ‘water.’

However, rejection of apriorism would be premature, given only what we’ve said so far. For we as yet really have no plausible account of the meaning of ‘water’, and no clear understanding of the sense in which it may be true that ‘water’ and other natural kind terms have objectual meanings.

The problem is that (6) and other definitions like it are clearly inadequate, for reasons that are independent of the Twin Earth case. It is obviously a logical possibility that XYZ, rather than H<sub>2</sub>O, should have been the colorless, odorless, thirst-quenching liquid found on Earth, and that no H<sub>2</sub>O exists on Earth at all. In this possible situation, stuff that is colorless, odorless, thirst-quenching liquid found on Earth would not be water, and stuff that is water would not be found on Earth. Thus the condition expressed by (6)’s *definiens* is neither necessary nor sufficient for application of the predicate ‘is water’.

In order to achieve a clear understanding of the problem posed for apriorism by the Twin Earth case, we first need a better account of the meaning of natural kind terms.

### 5. *Natural Kinds and Reference-Fixing*

The best explanation of the failure of such definitions as (5) and (6) seems to be found in the suggestion, made forcefully by both Putnam (1970, 1975) and Kripke (1972), that many general terms – such terms as ‘water’, ‘gold’, ‘aluminum’, ‘tiger’, and ‘lemon’ – are used simply to predicate membership in certain natural kinds, while the characteristics that we commonly associate with such terms are neither necessary nor sufficient for membership in the relevant kinds. One reason why these latter characteristics cannot be used to provide conditions for membership in natural kinds is that membership in a natural kind is an *essential* property of anything that is a member of that kind, while the characteristics in question are all non-essential, or contingent, properties.<sup>5</sup>

For example, to be a tiger is to belong to a certain natural kind, namely, the species *Felis tigris*. Nothing that is a tiger could exist without belonging to this kind, and necessarily, anything that belongs to this kind is a tiger. But the characteristics that we normally associate with tigers – such as those of being ferocious, tawny, black-striped, and four-legged – are all characteristics that non-tigers could possess and that tigers themselves could lack. Thus we cannot *define* the word ‘tiger’ in terms of these characteristics, for to be a tiger is to belong to a certain natural kind, and membership in this kind is not necessarily determined by possession of any of these contingent characteristics. On the other hand, the biological characteristics that *do* determine membership in the species *Felis tigris* cannot be used to define the word ‘tiger’ either, since the word had its ordinary meaning long before anyone knew what these biological characteristics were.

Natural kind terms such as ‘water’ and ‘tiger’ are used simply to predicate membership in certain natural kinds. Thus when we say of a sample of stuff  $x$  that it is water, we are predicating a certain relational property of  $x$ , namely the property of belonging to  $K$ , where  $K$  is the natural kind to which all samples of water belong. This is why

'water' and other natural kind terms cannot be defined in terms of the superficial contingent characteristics possessed by the instances of natural kinds. But if we are prevented from giving such definitions, then how can we construct a theory of meaning for 'water' and other such terms?

It is beginning to seem doubtful that natural kind terms even *have* definitions. Both Kripke (1972) and Putnam (1975) in fact suggest that the meanings of such words must be explained in a different way. Roughly, their idea is that a term like 'water' or 'tiger' functions as a proper name that refers directly to, or "rigidly designates," a given natural kind. On this view, such terms do not express the superficial contingent characteristics possessed by members of natural kinds. Rather the terms refer directly to the kinds themselves, while the contingent characteristics of the kinds' members serve instead to *fix the reference* of the terms in question. For example, we might fix the reference of 'tiger' relative to every possible world as being that biological kind to which certain ferocious, black-striped, cat-like animals *in fact* belong, that is, the kind to which such animals belong *in the actual world*. Similarly, we might fix the reference of 'water' relative to every possible world as being that natural kind to which liquid having such-and-such characteristics belongs in the actual world.

On this idea, the meaning of the word 'water' in English might be given by a reference-fixing rule of the following sort:

- (7) For any possible world  $w$  and entity  $x$ , 'water' is to refer to  $x$  relative to  $w$  if and only if: there is a natural kind  $K$  such that  $K$  exists in  $w$ ,  $x = K$ , and in the actual world, the colorless, odorless, thirst quenching liquid found on Earth belongs to  $K$ .

Let us call the proposal that 'water' has its meaning in English by virtue of the fact that English contains a reference-fixing rule like (7), "the reference-fixing proposal." This proposal has a significant advantage over proposals of definitions like (5) and (6).

Let us assume that a quantity of stuff  $x$  satisfies the predicate 'is water' in a possible world  $w$  if and only if in  $w$ ,  $x$  belongs to the kind to which 'water' refers in  $w$ . According to (7), 'water' refers to the

same kind relative to every possible world in which it has a referent at all, namely the kind  $K$  to which samples of water actually belong. Thus (7) has the consequence that for every possible world  $w$ , a quantity of stuff  $x$  satisfies the predicate 'is water' in  $w$  if and only if  $x$  belongs to  $K$  in  $w$ . Unlike (5) and (6), then, (7) has the advantage of being consistent with the idea that when we say of  $x$  that it is water we are saying simply that  $x$  belongs to  $K$ .

### 6. *A Problem with Referencing-Fixing*

This significant advantage of the reference-fixing proposal persuades me that it is on the right track, and that something close to it must be the correct account of natural kind terms.<sup>6</sup> However, the proposal is not quite correct as it stands, for it assumes that natural kind terms function like proper names that directly refer to natural kinds, and this assumption is counterintuitive in certain respects.

We want our theory of 'water' to have the consequence that the predicate 'is water' is used to predicate the relational property of belonging to  $K$ , where  $K$  is the natural kind to which samples of water in fact belong. The reference-fixing proposal seeks to achieve this result *via* the hypothesis that 'water' refers directly to  $K$ . But to use this hypothesis to get the result that 'is water' predicates the property of belonging to  $K$ , we apparently need to assume in addition that 'is' expresses the relation of *belonging to*. This in turn requires us to read a sentence of the form ' $x$  is water' such as

- (8) The stuff in my bathtub is water

as meaning the same as

- (9) \*The stuff in my bathtub belongs to water.

But (9) is ungrammatical, a fact that casts doubt on the hypothesis that in sentences like (8), 'water' functions as a genuine term that refers to a natural kind. Moreover, it seems clear that the copula 'is' never means 'belongs to.' For if it did, then it would be possible to say of a given object  $x$  that it belongs to a given kind  $K$  by saying something of the form ' $x$  is kind  $K$ '. But in sentences of this latter form, the 'is' can only be the 'is' of identity and so it cannot mean 'belongs to'.

Nathan Salmon has suggested that in predicates such as 'is water', the copula 'is' means 'is a sample of' (Salmon, 1981, p. 99). On this suggestion, (8) would mean

(10) The stuff in my bathtub is a sample of water.

Now (8) and (10) certainly do say the same thing. But in order for this fact to help with our present problem, we must also assume that in (10), the term 'water' is referring to a natural kind, and this assumption is implausible. For it seems clear that in (10), 'water' has the same meaning that it has in such sentences as

(11) My bathtub is full of water

and

(12) Jill poured water on Jack.

But in (11) and (12) 'water' does not refer to a natural kind. My bathtub, for instance, could not be full of an abstract natural kind. It could only be full of something physical, and in fact it never contains anything but wet physical stuff (water). But no abstract natural kind could be wet physical stuff. Similarly, (12) is true, not when Jill (impossibly) pours some natural kind on Jack, but only when Jill pours certain wet physical stuff on Jack. So in (11) and (12), 'water' does not refer to a natural kind, and since 'water' has the same meaning in (10) as it has in (11) and (12), it does not refer to a natural kind in (10) either. Rather, in such sentences as these, 'water' means 'stuff that is water', or more simply, it just means 'some water'.<sup>7</sup>

The reference-fixing proposal is *semantically* correct but *grammatically* wrong. The proposal is semantically correct because it correctly maintains that a sentence of the form 'x is water' predicates the relational property of belonging to *K*, where *K* is the natural kind to which water belongs. But the proposal gives the wrong grammatical account of how a sentence of this form comes to predicate this relational property. In particular, as we've seen, the fact that 'x is water' predicates the property of belonging to *K* cannot be correctly accounted for by the hypothesis that 'is' expresses the relation of *belonging to*, while 'water' refers to *K*.

How then can we account for the fact that 'is water' predicates a

relational property even though it is not relational in its grammatical form? I suggest that predicates like 'is water' get their meanings, not by means of rules that fix the reference of their component general terms, but instead by means of analogous rules that *fix the property* that is ascribed by means of the predicate. An example of the sort of rule I have in mind for 'is water' would be:<sup>8</sup>

- (13) For any  $\phi$ , if  $\phi$  is a token of 'is water', then for any property  $P$ ,  $\phi$  is to predicate  $P$  if and only if: there is a natural kind  $K$  such that in the actual world the colorless, odorless, thirst-quenching liquid found on Earth belongs to  $K$ , and  $P =$  the property of belonging to  $K$ .

Let us call the proposal that 'water' and other natural-kind terms get their meanings in English by virtue of the existence in English of property-fixing rules analogous to (13), the "property-fixing proposal." This proposal seems to have all the advantages of the reference-fixing proposal and none of its defects. In particular, the property-fixing proposal explains how it is possible for predicates containing natural kind terms to predicate relational properties even though such predicates are not relational in form, and even though they cannot be defined as equivalent to any other predicates of the language that are relational in form. In fact, one of the most interesting features of the property-fixing proposal is that it provides a useful alternative method of giving theories of meaning for predicates that resist the traditional sort of direct analysis in terms of necessary and sufficient conditions.

### PART III. THE TWIN EARTH PROBLEM AND ITS SOLUTION

#### 7. *Two Types of Meaning*

What I have been calling 'meaning', and what I intend to be the semantic concept with which apriorism is concerned, is perhaps best thought of as a particular species or type of meaning. This type of meaning is sometimes called *linguistic meaning*, since it is the kind of meaning that a word has in a particular language; that is, it is the kind of meaning that a word has by virtue of the semantic rules or



conventions that govern the word's use in a particular language. The property-fixing proposal is thus a theory about the linguistic meaning of natural kind terms, since it is a theory about the kind of semantic rule that governs such terms in particular languages.

But there is another concept of meaning that has loomed large in modern semantic theory. On this concept, the meaning of a word is identical with the specific semantic contribution that the word makes to the proposition expressed by a given sentence containing the word. I will call this sort of meaning, *propositional meaning*. The propositional meaning of a word is a semantic feature of the word that functionally determines which proposition is expressed by a sentence containing the word. This means that if a given sentence *S* contains a word *w* that has a certain propositional meaning, and a sentence *S'* results from *S* by substituting a word *w'* for *w* in *S*, then *S* and *S'* express the same proposition if and only if *w* and *w'* have the same propositional meaning. The propositional meaning of a whole sentence may simply be identified with the proposition that the sentence expresses.

The concept of propositional meaning is a close relative of Frege's (1892) concept of *sense*, since for Frege, the sense of a whole sentence is the proposition or thought it expresses, while the proposition expressed by a sentence is literally a function of the senses of its parts. We shall see shortly, however, that the two concepts are not precisely the same.

One of the most important discoveries that has been made in the philosophy of language is the discovery that linguistic meaning and propositional meaning are not the same thing. The distinction at the level of whole sentences was forcefully drawn by Strawson (1950) and later by Cartwright (1961), both of whom pointed out that a sentence with a single meaning in English may nevertheless express different propositions relative to different occasions of use. Thus, a sentence containing an indexical singular term, such as 'I am hungry' has a single linguistic meaning in English; but it expresses no particular proposition, since it expresses different propositions with perhaps different truth values when uttered either by different speakers or at different times.

Awareness of the same distinction at the level of particular words

has come more recently, due largely to the influence of David Kaplan's work on demonstratives and other indexical singular terms.<sup>9</sup> An indexical such as 'I' has a single linguistic meaning in English. But the propositional meaning of any indexical like 'I' must be allowed to vary from one context of use to another. Thus consider two tokens of the sentence 'I am hungry' that are uttered by different speakers at the same time. We have seen that these two tokens express different propositions. But to account for this difference, we must suppose that the two tokens of 'I' in the two sentence tokens make different semantic contributions, or in other words, have different propositional meanings. And since these two tokens of 'I' have the *same* linguistic meaning (the meaning of 'I' in English), we must conclude that the linguistic meaning of a singular term may be distinct from its propositional meaning.

But what is the propositional meaning of (a particular token of) a singular term like 'I'? Recent work in the theory of reference, particularly Kripke's (1972) work concerning the modal properties of sentences containing proper names, has tended to support the view that all singular terms – with the exception of definite descriptions – are genuine terms. Again, a "genuine term" is a term whose sole semantic contribution consists of its referent, a term, in other words, whose propositional meaning just *is* its referent.

If we identify the propositional meaning of a genuine term with its referent, then it is plausible to identify the linguistic meaning of such a term with a semantic rule that determines – or fixes – its referent.<sup>10</sup> For example, it is plausible to suppose that the linguistic meaning of the word 'I' is given by the reference rule:

- (14) For any  $\alpha$ , if  $\alpha$  is a token of 'I', then for any object  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $x$  is the speaker of  $\alpha$ .

It would seem then that *the linguistic meaning of a term determines its referent*. This doctrine stands in sharp contrast to Frege's view that *sense* determines reference. (Frege, 1892). For Frege, a term's sense must play two distinct semantic roles, since a term's sense is *both* the specific semantic contribution of the term to the proposition expressed – the term's propositional meaning – *and* the semantic feature of the term that determines its referent. But given that there

are genuine terms, we must distinguish the semantic features that play these two roles. For a genuine term's propositional meaning does not determine, but is identical with, its referent; and so the term's referent must be determined by something else – its linguistic meaning. From this point of view, Frege's concept of sense is an unfortunate conflation of two different concepts of meaning.<sup>11</sup>

It is plausible to suppose that the same kind of relation that holds between the linguistic and propositional meanings of genuine terms holds generally for all words and sentences. Thus I will say that, in general, *linguistic meaning determines propositional meaning*. If we apply this principle to natural kind terms, we should say that the linguistic meaning of such a term is given by a rule such as (13) that fixes (determines) the property that is predicated by use of the term, while the propositional meaning of the term is the property it is used to predicate. In the case of 'water', as we've seen, the property predicated is the purely relational property of belonging to *K*, where *K* is the natural kind to which samples of water in fact belong, and so this relational property is the propositional meaning of the word 'water'.

#### 8. *The True Significance of Twin Earth*

The view that most ordinary singular terms are genuine terms implies that there are *objectual* propositional meanings, that is, propositional meanings that essentially involve particular contingent objects. For instance, the propositional meaning of an ordinary proper name or indexical (relative to a given occasion of use) just is the ordinary contingent object that is the term's referent. Another kind of objectual propositional meaning is the singular proposition that is expressed by a sentence containing a genuine term whose referent is a contingent object. Since the singular proposition expressed by such a sentence is functionally determined by the referent of the genuine term it contains, it surely seems that the singular proposition in question could not exist unless this referent existed. Thus some singular propositions seem to essentially involve particular contingent objects.

The main evidence in favor of objectual propositional meanings is the evidence in favor of genuine terms, such as the considerations

adduced by Kripke (1972) concerning the modal properties of sentences containing proper names. As we saw earlier, Kripke's (1972) and Putnam's (1970, 1975) modal considerations regarding natural kind terms support a similar result concerning the propositional meanings of such terms, since as we've seen, these considerations support the conclusion that natural kind terms are used to predicate purely relational properties that essentially involve certain natural kinds.

But the Twin Earth example supports a different and even more startling consequence. This is the consequence that *there are objectual linguistic meanings*. Recall our earlier example of a property-fixing rule:

- (13) For any  $\phi$ , if  $\phi$  is a token of 'is water', then for any property  $P$ ,  $\phi$  is to predicate  $P$  if and only if: there is a natural kind  $K$  such that in the actual world, the colorless, odorless, thirst-quenching liquid found on Earth belongs to  $K$ , and  $P =$  the property of belonging to  $K$ .

Since (13) contains an occurrence of a genuine term – the proper name 'Earth' – the rule expressed by (13) is a function of the term's referent, the planet Earth. This rule, which is the propositional meaning of the sentence (13), is therefore itself an objectual propositional meaning that is contingent for its existence on the existence of the planet Earth. But if the word 'water' were governed in English by the semantic rule (13), this fact would be the same as the fact that 'water' has a given linguistic meaning in English. Thus, since the rule (13) is contingent for its existence on the planet Earth, it would follow that the word 'water' could not have the linguistic meaning that it in fact has in English, unless the planet Earth existed. In this way, the linguistic meaning of 'water' would be objectual – would essentially involve a particular contingent object – if 'water' were governed in English by an objectual rule like (13).

And the Twin Earth case seems to show that 'water' must be governed by an objectual rule analogous to (13). Deletion of the proper name 'Earth' in (13) would leave us with a purely general rule that could not be distinguished from the rule that would be followed by our doppelgängers on Twin Earth. So unless the rule that we fol-

low in using 'water' essentially involves some object that is distinct from any object that is involved in the similar rule that would be followed by our doppelgängers on Twin Earth, the linguistic meaning of 'water' on Earth would be indistinguishable from its linguistic meaning on Twin Earth. But then, the properties (if any) predicated by use of 'is water' on Earth and Twin Earth, as well as the extensions of the predicate on both planets, would also be indistinguishable. To avoid this problem, it seems, we must assume that 'water' is governed in English by a property-fixing rule that is objectual. Thus, we must assume that the linguistic meanings of 'water' and other natural kind terms are also objectual.

This important consequence of the Twin Earth example regarding linguistic meaning has been obscured, in part because discussions of the example usually indicate no awareness of the distinction between the two types of meaning. Putnam's own discussion, for instance, relies on the concept of a word's *intension*, but since this concept is just the same as Frege's concept of sense, it also conflates propositional and linguistic meaning.

We have seen that there are both objectual propositional meanings and objectual linguistic meanings. But it is only the existence of objectual linguistic meanings that is of direct relevance to apriorism. For apriorism is plausible only as a principle about linguistic meaning. This is because the propositional meaning of a word or sentence is typically determined in part by the word's or sentence's linguistic meaning, and in part by features of the context of utterance and other contingent facts about the world. But such features and facts are of course not the sort of thing that one can know a priori. Good examples of this are provided by genuine terms and sentences that contain them. The propositional meaning of a genuine term is its referent, but one cannot in general know a priori which object is a term's referent, or even whether the term has a referent at all. And since the singular proposition expressed by a sentence containing such a term is a function of the term's referent, one cannot in general know a priori which proposition the sentence expresses, or even whether the sentence expresses a proposition at all.

It would thus be a mistake to assume that propositional meanings are knowable a priori. But for the reasons discussed earlier in Part I

the situation is different with respect to linguistic meaning. For again, a speaker's linguistic competence regarding a word seems to consist of that person's having internalized the semantic rule of his language by virtue of which the word has its linguistic meaning in that language. And this internalizing of a word's meaning would seem to be one and the same as the process whereby a speaker acquires those linguistic propensities that should enable the speaker to know a priori what she or he means by the word.

But the Twin Earth example shows that there are objectual linguistic meanings, and this fact is difficult to reconcile with apriorism. Consider again the property-fixing rule expressed by (13). This rule directly and essentially involves the planet Earth, so that if Earth did not exist the rule expressed by (13) would also not exist. And since no one can know a priori that Earth exists, no one can know a priori that the rule (13) exists. Thus it would be impossible to know a priori that the word 'water' has the meaning given by the rule (13). It would even be impossible for a person to know a priori that this is what he or she means by 'water', since knowledge of this latter fact would also imply that the planet Earth exists. So if the word 'water' were governed in English by the rule (13), then apriorism would be false. My problem, then, is to find an objectual property-fixing rule analogous to (13) that may plausibly be taken to give the meaning of 'water' in English, but that unlike (13) does not essentially involve a contingent object whose existence cannot be known a priori.

### 9. *Solution to the Twin Earth Problem*

As we saw earlier in Section 1, there is no shortage of contingent objects whose existence can be known a priori. Thus, I know a priori that I exist, that the present time exists, and that various of my present mental experiences and states exist. However, these and other common examples of objects that are knowable a priori are *private* objects, in the sense that their existence can only be known a priori by a single person. As such, objects like this are poor candidates for constituents of the public meanings of natural kind terms such as 'water'.

For instance, consider the result of replacing the name 'Earth' in (13) by the description 'the planet that *I* inhabit'. The rule that I

would express by this result is a rule whose existence I know a priori, since I myself am the only contingent object whose existence is entailed by this rule, and I know a priori that I exist. But the hypothesis that speakers of English follow rules of this kind in using 'water' is quite implausible, since it implies that each speaker follows his own private rule for 'water' and that consequently 'water' has no common public meaning in English.<sup>12</sup>

The difficulty for apriorism may seem insurmountable. For it seems that the only objects that can serve as constituents of meanings that are knowable a priori are private objects that are unfit to serve as constituents of meanings in a public language. Still, I think that there is a kind of object that can serve as a constituent of public meanings and that in a relevant sense is also knowable a priori.

Whenever I try to explain what I mean by 'water' in such a way as to distinguish what I mean from what my counterpart on Twin Earth would mean, I find myself saying such things as "the colorless, odorless, thirst-quenching liquid found on *our* planet" or "the colorless, odorless, thirst-quenching liquid that *we* have experienced." Here of course the emphasized 'our' and 'we' are intended to distinguish *us* (the inhabitants of Earth) from *them*, (the inhabitants of Twin Earth). Perhaps, then, the group consisting of Earth's inhabitants could be used to distinguish the meaning of 'water' as we use it from the meaning that 'water' would have on Twin Earth.

Perhaps, for instance, the property-fixing rule that gives the meaning of 'water' in English could be expressed as follows:

- (15) For any  $\phi$ , if  $\phi$  is a token of 'is water', then for any property  $P$ ,  $\phi$  is to predicate  $P$  if and only if: there is a natural kind  $K$  such that in the actual world, the colorless, odorless, thirst-quenching liquid that *we* have experienced belongs to  $K$ , and  $P =$  the property of belonging to  $K$ .

An English speaker on Earth, I assume, could use the pronoun 'we' in (15) to demonstratively pick out a group consisting of himself and other inhabitants of Earth, so that in this speaker's mouth, (15) would express an objectual rule that essentially involves that group. Such a speaker's doppelgänger on Twin Earth, by contrast, would use 'we' in (15) to demonstratively pick out the group consisting of himself and

other inhabitants of Twin Earth, so that the rule expressed would be a different rule that essentially involves a different group.

Now of course in order for this proposal to help apriorism, it must be possible for speakers of English in general to know a priori that the relevant group exists. But is this really possible, and if so, how?

The indexical 'we' always refers, relative to a given occasion of use, to a group that includes at least the speaker and, depending on the context, other individuals that the speaker has in mind and that are related to the speaker in some way. Typically, a user of 'we' will pick out the group referred to from his own perspective, using himself as a reference point. There are in fact various methods of this kind that each speaker could use to pick out the whole of humanity.

For instance, one might specify "that group consisting of myself and all those who both belong to the same natural kind as myself and are genetically related to me by descent." Or alternatively, one might specify "that group consisting of myself and all those who both belong to the same natural kind as myself and who inhabit the same planet as I do." There are no doubt many different possible methods like these that could be used by individual speakers to pick out humanity as a whole, and it is not part of my view that there is any *single* method that all speakers would use. To defend apriorism, I need to show only that there are some methods that speakers could use to specify humanity as a whole, and that speakers can know a priori that these methods succeed in specifying an existing group.

The two methods just described both satisfy this last condition. Thus, I know a priori that there exists a group consisting of myself and all those who both belong to the same natural kind as myself and are genetically related to me by descent. This knowledge is of course parasitic upon my a priori knowledge that I exist. Should it turn out that I live in a solipsistic world in which no one other than myself exists, then the group I have specified is a group that has myself as its sole member. Nevertheless, I know a priori that I have succeeded in specifying a group. What I cannot know a priori is whether this group has more than one member.

Now suppose that after careful consideration of my own linguistic intuitions and dispositions, I come to know a priori that I intend my uses of the predicate 'is water' to be subject to a certain semantic rule.



The rule in question, suppose, is the one I would express by use of (15), with the referent of 'we' determined by the condition "group consisting of myself and all who both belong to the same natural kind as myself and are genetically related to me by descent." I can specify this rule knowing a priori that this specification is satisfied, and so I know a priori that this rule exists. If I live in a solipsistic world, then the rule has as a constituent a group that has myself as its sole member. But if, as I suspect, I do not live in a solipsistic world, then the constituent referred to by my use of 'we' is a huge group that consists of all the humans who inhabit Earth. But the group would not include the inhabitants of Twin Earth; those humans, if there were such, would not be genetically related to me.

Notice that if the world is not solipsistic and other English speakers mean by 'water' what I do, then we all end up following the same semantic rule, namely, the rule expressed by (15), where 'we' refers to the inhabitants of Earth. Thus, in these circumstances, 'water' has a common public meaning. But at the same time, each speaker who correctly understands this public meaning can, by following the procedure I've described, know a priori what that meaning is. Indeed, if I am right that the rule (15) gives the meaning of 'water' in English, then I myself am an example of someone who has come to know a priori what this meaning is.

#### 10. *A Serious Objection*

The group that serves as a constituent of the objectual meaning of 'water' and other natural kind terms must not be identified with a *class*, if my solution to the Twin Earth problem is to work. Surely, in other possible worlds in which there are slightly fewer or more humans on Earth than there actually are, we could still use 'water' with the same meaning as we actually do. But in each of these other worlds the class of humans on Earth is distinct from the actual class of humans on Earth. So if our meaning of 'water' is to remain the same in these other worlds the group of genetically related humans mentioned in the rule that governs 'water' cannot be a class. Instead, the group in question must be an abstract entity that, unlike a class, can persist in existence through changes in its membership. In this

respect, the group would resemble such things as the New York Yankees and the U.S. Supreme Court.<sup>13</sup> But this consequence seems unobjectionable. For surely, the human race is in fact a group that persists in existence through radical changes in its membership.

A related consideration poses a more serious difficulty. It seems to follow from my solution that in a solipsistic world in which I am the only human, the meaning of 'water' as I use it would have to be different from its meaning in the actual world. For surely, the group that consists solely of myself and the group that consists of all the (actual) humans on Earth must be distinct groups. Hence the rule of the form (15) that I would intend to follow in the solipsistic world is different from the rule that I intend to follow in the actual world, since the two rules contain different groups as constituents. So in the solipsistic world I would use 'water' with a different meaning than in the actual world. But of course I cannot know a priori that my world is *not* solipsistic. Hence I cannot know a priori *which* of the two possible meanings of 'water' I actually intend my uses of 'water' to have. But then, or so the objection goes, I cannot be said to be capable of knowing a priori what meaning I actually intend my uses of 'water' to have.

Now I am not entirely certain that the relevant groups in the two different worlds in question really would be distinct. After all, since the groups are not classes, their identities are not determined solely by their membership. So perhaps a case could be made that the group in the solipsistic world that consists solely of myself is one and the same as the group in the actual world that consists of all humanity including myself.<sup>14</sup>

However, since I am uncertain about this point, I do not want my solution to the Twin Earth problem to depend on it. So for the sake of argument, I will grant that the relevant groups are distinct. I will also grant that I cannot know a priori which of the two groups are constituents of the meaning that I intend my uses of 'water' to have, and because of this, I also cannot know a priori which of the two possible meanings it is that I intend my uses of 'water' to have. However, I deny that it follows from this that I cannot know a priori what I actually intend 'water' to mean (that is, what I mean by 'water' in the actual world).

For brevity, let  $R$  be the relation by which I pick out humanity as “the group consisting of myself and all who bear  $R$  to me.” Also, assume for the sake of argument that I am right in my claim that (15) accurately expresses what I mean by ‘water’. Then in both the solipsistic world and the actual world, I know a priori what I mean by ‘water’ in the sense that in both worlds, the following is true:

- (16) I know a priori that there exists a group consisting of myself and all who bear  $R$  to me, and I know a priori that I intend my uses of ‘is water’ to be subject to the following rule: tokens of ‘is water’ are to predicate a property  $P$  if and only if  $P$  is the property of belonging to  $K$ , where  $K$  is the natural kind to which the colorless, odorless, thirst-quenching liquid experienced by *that group* belongs.

(16) is an instance of a general type of sentence that was first studied by Geach (1967). A simpler instance of the type in question is provided by this example of Geach’s:

- (17) Nob assumes that just one witch blighted Bob’s mare, and Nob wonders whether *she* (that same witch) killed Cob’s sow.

I have proposed elsewhere (McKinsey, 1986) that in contexts like these, the pronoun or demonstrative (‘she’ in (17), ‘that group’ in (16)) has the meaning of what Evans (1977) has called an “E-type pronoun.” An E-type pronoun is an indexical genuine term whose reference is fixed by the definite description that is recoverable from the clause governed by the pronoun’s quantifier antecedent. My proposal in (1986) was that when such a term occurs in the scope of an attitude-operator, as in (16) and (17), the term’s use represents a mental act whose objective reference is fixed by the agent’s descriptive assumption, in a manner analogous to the way an E-type pronoun’s reference is fixed by its antecedent.

Given this understanding, then, and given that the first conjunct of (16) is true, the term ‘that group’ in (16) is a genuine term that refers to the group consisting of myself and all who bear  $R$  to me. Hence, the clause in the scope of ‘intends that’ in (16) succeeds in expressing a semantic rule that has the relevant group as a constituent. And

according to the second conjunct of (16), I know a priori that I intend my tokens of 'is water' to be subject to this semantic rule. In other words, I know a priori that I intend my tokens of 'is water' to have a certain linguistic meaning.

Notice that given our assumptions, (16) is true in both the actual world and the solipsistic world. This holds even if we grant that in these two worlds the term 'that group' in (16) would refer to distinct groups, so that the meanings I intend 'water' to have in the two worlds would also be distinct. Since (16) is true in both worlds, I know a priori in each world that I intend 'water' to have a certain meaning, although in neither world do I know a priori *which* of the two possible meanings I have in mind.

In each world, I know a priori only that I have in mind the group consisting of myself and all who bear *R* to me. But in neither world do I know a priori *which* group this is, since I do not know a priori whether or not the group in question has anyone in it besides myself. But it is important to see that my failure to know a priori which group I have in mind is merely a failure of *de re* knowledge on my part. Consequently, my failure to know a priori which meaning of 'water' I have in mind is also a failure of *de re* knowledge. But this failure does not affect the a priori *de dicto* knowledge that is ascribed to me in (16). Since (16), by assumption, is true, I know a priori (*de dicto*) that I mean a certain thing by 'water', even though I do not know a priori (*de re*) which meaning of 'water' it is that I have in mind.

### 11. *Concluding Remarks*

It is worth noting an important historical precedent of the view I have tried to defend in this paper. The precedent is found in Russell's well known view that the referents of genuine terms, and in fact all the constituents of the propositions that one can understand, must be objects with which one is acquainted. Russell called this principle "The Principle of Acquaintance." (Russell, 1912, Chapter Five.) For Russell, we are acquainted only with things that we are mentally presented with, things such as our own mental acts and states,

abstract entities like propositions and universals, and the constituents of sense perception, which Russell called “sense-data.”

Russell’s objects of acquaintance, of course, are precisely the objects whose existence is knowable a priori, and so we can see that his Principle of Acquaintance is the result of a kind of misguided apriorism, an apriorism that Russell shared in common with Frege. The apriorism is misguided, since Russell applied it to *propositional* meanings, whereas as we’ve seen, apriorism is plausible only as a principle about *linguistic* meanings. So Russell endorsed his Principle of Acquaintance because like Frege, he also conflated the two types of meaning.

However, because of the Twin Earth example, something very much like Russell’s Principle of Acquaintance has come back to haunt us. For the example shows that some linguistic meanings have objectual constituents, and apriorism regarding linguistic meaning implies that these constituents must be knowable a priori. The view I have defended, therefore, is the correct counterpart concerning linguistic meaning of Russell’s incorrect principle about propositional meaning.

I have tried to defend apriorism primarily because I believe it is a principle whose truth is required by our concepts of what it is to speak and understand the words of a language. But it is also worth re-emphasizing that apriorism seems to be a presupposition of much of philosophy as it is done today. Consequently, a philosopher who uses examples like the Twin Earth case to attack apriorism would seem to be guilty of a kind of pragmatic inconsistency. Such a philosopher appeals to his and others’ linguistic intuitions concerning the meaning of ‘water’ and other natural kind terms in order to argue that apriorism is false. But if apriorism is false, then it certainly seems that one’s linguistic intuitions should just be *irrelevant* to settling any issues about meaning. So a philosopher who believes that apriorism is false and who at the same time keeps appealing to linguistic intuitions as a way of settling philosophical disputes certainly owes the rest of us some explanation of how it is possible to consistently do this. Of course, if my defense of apriorism is correct, no such explanation is necessary.

## NOTES

<sup>1</sup> See Putnam (1975), pp. 136–137. I have added the qualification “contingent” to Putnam’s characterization of the Fregean doctrine. Without the qualification, the doctrine is trivially false, since there are presumably an infinite number of necessarily existing objects (the natural numbers, say) whose existence is entailed by the existence of any psychological state.

<sup>2</sup> My thanks to Alvin Plantinga for bringing this problem to my attention. Kripke (1972) briefly discusses the problem, pp. 260–261, as does Plantinga (1974), pp. 7–8.

<sup>3</sup> On the distinction between word-meaning and speaker-meaning, see for instance Ziff (1967) and Grice (1968).

<sup>4</sup> Thus apriorism in the philosophy of language (in the sense of (4) below) is a corollary of apriorism in the philosophy of mind. The idea that one can have a priori knowledge of one’s own mental states is of course at least as old as Descartes. I have argued elsewhere (McKinsey, 1978) that this compelling Cartesian intuition is in conflict with certain recently fashionable views concerning reference and intentionality, including certain forms of the “causal theory of reference.”

<sup>5</sup> A property  $P$  is an *essential* property of an object  $x$  if and only if:  $x$  has  $P$ , and it is logically impossible that  $x$  exists and yet fails to have  $P$ . A property  $P$  is a *contingent* property of an object  $x$  if and only if  $x$  has  $P$ , but  $P$  is not an essential property of  $x$ .

<sup>6</sup> Both Kripke and Putnam have also suggested a slightly different model of reference-fixing on which the reference of a natural kind term is fixed by use of paradigmatic instances (Kripke, 1972, p. 319; Putnam, 1975, p. 148.) On this model, the meaning of ‘water’ would be given by a rule of the following sort:

- (i) For any possible world  $w$  and entity  $x$ , ‘water’ is to refer to  $x$  relative to  $w$  if and only if: there is a kind  $K$  such that  $K$  exists in  $w$ ,  $x=K$ , and *this liquid* belongs to  $K$  in the actual world,

where ‘this liquid’ demonstratively picks out a particular sample of water in the actual world. Though this proposal has some of the same advantages as that of (7) it does not provide an adequate theory of the meaning of ‘water’. For the meaning of ‘water’ in English is something *public*, something that is understood or grasped in common by past and present speakers of English. But a rule of the form (i) essentially involves a specific sample of water that would be known by only a few speakers at most, and so no such rule can give expression to the shared public meaning of the word ‘water’. The proposal of (7) of course avoids this problem since instead of relying on idiosyncratic samples, it relies on characteristics that are commonly associated with the word ‘water’ by speakers of English.

<sup>7</sup> The above discussion suggests the sort of view of ‘water’ and other mass nouns that Burge (1972) has proposed. On this view, mass nouns are fundamentally predicates, as opposed to individual constants. The basic meaning of a mass noun like ‘water’ is the meaning that it has in a predicate of the form ‘is N’ (‘is water’), while the noun’s other meanings are definable in terms of this basic one. For instance, as Burge suggests (p. 277), the logical form of (8) would be

- (8a) Water (the stuff in my bathtub),

while the form of (11), in which ‘water’ means ‘some water’, would be

- (11a)  $(\exists x) (\text{Water}(x) \ \& \ \text{Full-of}(\text{my bathtub}, x))$ .

Similarly, we can treat ‘water’ in sentences like

- (i) Water is wet

as meaning 'all water', so that this sentence would mean

(ii)  $(\forall x)(\text{Water}(x) \rightarrow \text{Wet}(x))$

<sup>8</sup> In formulating (13), I have eliminated relativization to possible worlds, since I assume there is just one property that a predicate with a given meaning could be used to predicate, so that which property is predicated does not vary from world to world. Of course one could relativize predication to possible worlds, but since a predicate will always predicate the same property relative to every world, the complication is unnecessary.

<sup>9</sup> See for instance Kaplan (1979). What I am calling 'linguistic meaning' and 'propositional meaning', Kaplan calls 'character' and 'content', respectively. The most thorough theoretical treatment to date of the distinction between the two kinds of meaning is found in Pollock (1982). Pollock uses the terms 'meaning' and 'sense' for linguistic meaning and propositional meaning, respectively.

<sup>10</sup> I have made this proposal in greater detail, and with application to proper names, in McKinsey (1984).

<sup>11</sup> This criticism of Frege is similar to one made by Kripke (1972, p. 277), who also accuses Frege of conflating two concepts under the term 'sense'. But Kripke does not explicitly associate the way a term's reference is determined with its linguistic meaning.

<sup>12</sup> The solution that Searle has proposed to the Twin Earth Problem has this kind of defect, in my opinion. (Searle, 1983, pp. 207–208).

<sup>13</sup> For a discussion of such entities and their difference from classes, see Sharvy (1968).

<sup>14</sup> This possible response was suggested to me by Lawrence Lombard, though he does not necessarily endorse it.

#### REFERENCES

- Burge, Tyler: 1972, 'Truth and Mass Terms', *Journal of Philosophy* 69, 263–282.
- Cartwright, Richard: 1961, 'Propositions', in R. J. Butler (ed.), *Analytic Philosophy*, Vol. 1, Basil Blackwell, Oxford.
- Evans, Gareth: 1977, 'Pronouns, Quantifiers, and Relative Clauses (I)', *Canadian Journal of Philosophy* 7, 467–536.
- Frege, Gottlob: 1892, 'On Sense and Reference', in *Translations from the Writings of Gottlob Frege*, Basil Blackwell, Oxford, 1966.
- Geach, P. T.: 1967, 'Intentional Identity', *Journal of Philosophy* 64, 627–632.
- Grice, H. P.: 1969, 'Utterer's Meaning and Intentions', *Philosophical Review* 78, 147–177.
- Kaplan, David: 1979, 'On the Logic of Demonstratives', in P. French, T. Uehling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language*, University of Minnesota Press, Minneapolis, pp. 401–412.
- Kripke, Saul: 1972, 'Naming and Necessity', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, D. Reidel, Dordrecht, pp. 253–355.
- McKinsey, Michael: 1978, 'Names and Intentionality', *The Philosophical Review* 87, 171–200.
- McKinsey, Michael: 1984, 'Causality and the Paradox of Names', *Midwest Studies in Philosophy* 9, 491–515.
- McKinsey, Michael: 1986, 'Mental Anaphora', *Synthese* 66, 159–175.
- Plantinga, Alvin: 1974, *The Nature of Necessity*, Clarendon Press, Oxford.
- Pollock, John L.: 1982, *Language and Thought*, Princeton University Press, Princeton, New Jersey.

- Putnam, Hilary: 1970, 'Is Semantics Possible?', in S. Schwartz (ed.), *Naming, Necessity, and Natural Kinds*, Cornell University Press, Ithaca, 1977.
- Putnam, Hilary: 1975, 'The Meaning of "Meaning"', in K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science VII: Language, Mind, and Knowledge*, University of Minnesota Press, Minneapolis, pp. 131–193.
- Russell, Bertrand: 1912, *The Problems of Philosophy*, Oxford University Press, New York, 1959.
- Salmon, Nathan U.: 1981, *Reference and Essence*, Princeton University Press, Princeton, New Jersey.
- Searle, John R.: 1983, *Intentionality*, Cambridge University Press, Cambridge.
- Sharvy, Richard: 1968, 'Why a Class Can't Change Its Members', *Noûs* 2, 303–314.
- Strawson, P. F.: 1950, 'On Referring', *Mind* 59, 320–344.
- Ziff, Paul: 1967, 'On H. P. Grice's Account of Meaning', *Analysis* 28, 1–8.

*Department of Philosophy,*  
*Wayne State University,*  
*Detroit, MI 48202,*  
*U.S.A.*