II

Explanation and reference*

I. General significance of the topic

In this paper I try to contrast realist theories of meaning with what may be called 'idealist' theories of meaning. But a word of explanation is clearly in order.

There is no Marxist 'theory of meaning' but there are a series of remarks on the correspondence between concepts and things, on concepts, and on the impossibility of *a priori* knowledge in the writings of Engels (cf. Engels, 1959) which clearly bear on problems of meaning and reference. In particular, there is a passage† in which Engels makes the point that a concept may contain elements which are not correct. A contemporary scientific characterization of fish would include, Engels says, such properties as life under water and breathing through gills; yet lungfish and other anomalous species which lack these properties are classified as fish for scientific purposes. And Engels argues, I think correctly, that to stick to the letter of the 'definition' in applying the concept *fish* would be bad science. In short, Engels contends that:

(1) Our scientific conception (I would say 'stereotype') of a fish includes the property 'breathing through gills', but
(2) 'All fish breath through gills' is not true! (and, *a fortiori*, not analytic).

I do not wish to ascribe to Engels an anachronistic sophistication about contemporary logical issues, but without doing this it is fair to say on the basis of this argument that Engels *rejects* the model according to which such a concept as *fish* provides anything like analytically necessary and sufficient conditions for membership in a natural kind. Two further points are of importance: (1) The fact that the concept 'natural kind *all* of whose members live under water, breath through gills, etc.' does not

* First published in G. Pearce and P. Maynard (eds.) *Conceptual Change* (Dordrecht-Reidel 1973) 199–221.

† In a letter written to Conrad Schmidt in 1895; cf. Marx (1942), pp. 527–30. My agreement is with Engels' realism, not his 'dialectical materialism'.

strictly fit the natural kind Fish does not mean that the concept does not *correspond* to the natural kind Fish. As Engels puts it, the concept is not exactly correct (as a description of the corresponding natural kind) but that does not make it a *fiction*. (2) The concept is continually changing as a result of the impact of scientific discoveries, but that does not mean that it ceases to correspond to the same natural kind (which is itself, of course, also changing). Again, without attributing to Engels a sophisticated theory of meaning and reference, it is fair, I think, to restate the essential gist of these two points in the following way: concepts which are not strictly true of anything may yet refer to something; and concepts in different theories may refer to the same thing. Of these two points, the second is obvious for most realists; with a few possible exceptions (e.g. Paul Feyerabend), realists have held that there are successive scientific theories about the *same* things: about heat, about electricity, about electrons, and so forth; and this involves treating such terms as 'electricity' as *trans-theoretical* terms, as Dudley Shapere has called them (cf. Shapere, 1969), i.e. as terms that have the same reference in different theories. The first point is more controversial; the idea that concepts provide necessary and sufficient conditions for class membership has often been attacked but, nonetheless, constantly reappears. Without it, however, the other point is moot. Bohr assumed in 1911 that there are (at every time) numbers p and q such that the (one dimensional) position of a particle is q and the (one dimensional) momentum is p; if this was part of the meaning of 'particle' for Bohr, and in addition, 'part of the meaning' means 'necessary condition for membership in the extension of the term', then electrons are *not* particles in Bohr's sense, and, indeed, there are *no* particles 'in Bohr's sense'. (And no 'electrons' in Bohr's sense of 'electron', etc.) None of the terms in Bohr's 1911 theory referred! It follows on this account that we cannot say that present electron theory is a better theory of the same particles that Bohr was referring to. I take it that this is the line of thinking that Paul Feyerabend represents. On an account like Shapere's, however, Bohr would have been referring to electrons when he used the word 'electron', notwithstanding the fact that some of his beliefs about electrons were mistaken, and *we* are referring to those same particles notwithstanding the fact that some of our beliefs – even beliefs included in our scientific 'definition' of the term 'electron' – may very likely turn out to be equally mistaken. This seems right to me. The main technical contribution of this paper will be a sketch of a theory of meaning which supports Shapere's insights.

An 'idealist' theory of meaning, as I am using the term, might go like this (in its simplest form): the meaning of such a sentence as 'electrons

exist' is a function of certain *predictions* that can be derived from it (in a pure idealist theory, these would have to be predictions about *sensations*); these predictions are clearly a function of the *theory* in which the sentence occurs; thus 'electrons exist' has no meaning apart from this, that or the other theory, and it has a different meaning in different theories.

The question of 'reference' is a harder one for an idealist: the essence of idealism is to view scientific theories and concepts as instruments for predicting sensations and not as representatives of real things and magnitudes. But a sophisticated idealist is likely to say that the question of reference is 'trivial':† if one has a scientific language L containing the term 'electron', then one can certainly construct a metalanguage ML over it *à la* Tarski, and define 'reference' in such a way that '"electron" refers to electrons' is a trivial theorem. But if different scientific theories T_1 and T_2 are associated with different formal languages L_1 and L_2 (as they must be if the words have different meanings in T_1 and T_2), then they will be associated with different *meta*-languages ML_1 and ML_2. In ML_1 we can say '"electron" refers to electrons', meaning that 'electron' in the sense of T_1 refers to electrons *in the sense of* T_1, and in ML_2 we can say '"electron" refers to electrons' meaning that 'electron' in the sense of T_2 refers to electrons *in the sense of* T_2; but there is no ML in which we can even express the statement that 'electron' refers to the same entities in T_1 and T_2 – or, at least, no prescription for constructing such an ML has been provided by Positivist philosophers of science. In short, just as the idealist regards 'electron' as *theory dependent*, so does he regard the semantical notions of reference and truth as theory dependent; just as the realist regards 'electron' as *trans-theoretical*, so does he regard truth and reference as trans-theoretical.

II. The meaning of physical magnitude terms

A. A causal account of meaning

My purpose here is to sketch an account of the meaning of physical magnitude terms (e.g. 'temperature', 'electrical charge'); not an account of meaning in general, although I will try to indicate similarities between what is said here about these terms and what Kripke has said about proper names and what I have said elsewhere about natural kind words. (Kripke's work has come to me second hand; even so, I owe him a large debt for suggesting the idea of causal chains as the mechanism of reference.)

On a traditional view, any term has an intension and an extension.

† See, for example, the discussion by Hempel (1965), pp. 217–18. A contrasting view is sketched in chapter 13, volume 1 of these papers.

'Knowing the meaning' is having knowledge of the intension; what it is to 'know' an intension (construed, usually, as an abstract entity of some kind) is never explained. The extension of the color term 'red', for example, is the class of red things; the intension, according to Carnap, is the property Red. Carnap spoke of 'grasping' the intension of terms; what it would be to 'grasp' the property Red was never explained; probably Carnap would have equated it with knowing how to verify sentences of the form 'x is red', but this comes from his theory of knowledge, not his writings on semantics. In any case, understanding words is a matter of having knowledge. Full linguistic competence in connection with a word may require more knowledge than just the intension; for example, syntactical knowledge, knowledge of cooccurrence regularities, etc.; but linguistic competence, like understanding, is a matter of *knowledge* – not necessarily explicit knowledge – knowledge in the wide sense, implicit as well as explicit, 'knowing how' as well as 'knowing that', skills and abilities as well as facts, but all *knowledge* none the less.

According to the theory I shall present this is fundamentally wrong. Linguistic competence and understanding are not just *knowledge*. To have linguistic competence in connection with a term it is not sufficient, in general, to have the full battery of usual linguistic knowledge and skills; one must, in addition, be in the right sort of relationship to certain distinguished situations (normally, though not necessarily, situations in which the *referent* of the term is present). It is for this reason that this sort of theory is called a 'causal theory' of meaning.

Coming to physical magnitude terms, what every user of the term 'electricity' knows is that electricity is a magnitude of some sort – and, in fact, not even that: electricity was thought at one time to possibly be a sort of substance, and so was heat. At any rate, speakers know that 'electricity' and 'heat' are putative physical *quantities* – capable of more and less, and capable of location. (I do not think that even these statements are *analytic*, but I think they have a kind of *linguistic* association with the terms in question.) In a developed semantic theory one might introduce a special semantic marker, e.g. 'physical quantity', for terms of this sort. I cannot, however, think of anything that *every* user of the term 'electricity' *has* to know except that electricity is (associated with the notion of being) a physical magnitude of some sort, and, possibly, that 'electricity' (or electrical charge or charges) is capable of flow or motion. Benjamin Franklin knew that 'electricity' was manifested in the form of sparks and lightning bolts; someone else might know about currents and electromagnets; someone else might know about atoms consisting of positively and negatively charged particles. They could all

use the term 'electricity' without there being a discernible 'intension' that they all share. I want to suggest that what they do have in common is this: that each of them is connected by a certain kind of causal chain to a situation in which a *description* of electricity is given, and generally a *causal* description – that is, one which singles out electricity as *the* physical magnitude *responsible* for certain effects in a certain way.

Thus, suppose I were standing next to Ben Franklin as he performed his famous experiment. Suppose he told me that 'electricity' is a physical quantity which behaves in certain respects like a liquid (if he were a mathematician he might say 'obeys an equation of continuity'); that it collects in clouds, and then, when a critical point of some kind is reached, a large quantity flows from the cloud to the earth in the form of a lightning bolt; that it runs along (or perhaps 'through') his metal kite string; etc. He would have given me an *approximately correct definite description* of a physical magnitude. I could now use the term 'electricity' myself. Let us call this event – my acquiring the ability to use the term 'electricity' in this way – an *introducing event*. It is clear that each of my later uses will be causally connected to this introducing event, as long as those uses exemplify the ability I acquired in that introducing event. Even if I use the term so often that I forgot when I first learned it, the intention to refer to the same magnitude that I referred to in the past by using the word links my present use to those earlier uses, and indeed the word's being in my present vocabulary at all is a causal product of earlier events – ultimately of the introducing event. If I teach the word to someone else by telling him that the word 'electricity' is the name of a physical magnitude, and by telling him certain facts about it which do not constitute a causal description – e.g. I might tell him that like charges repel and unlike charges attract, and that atoms consist of a nucleus with one kind of charge surrounded by satellite electrons with the opposite kind of charge – even if the facts I tell him do not constitute a definite description of any kind, let alone a causal description – still, the word's being in his vocabulary will be causally linked to its being in my vocabulary, and hence, ultimately, to an introducing event.

I said before that different speakers use the word 'electricity' without there being a discernible 'intension' that they all share. If an 'intension' is anything like a necessary and sufficient condition, then I think that this is right. But it does not follow that there are no ideas about electricity which are in some way linguistically associated with the word. Just as the idea that tigers are striped is linguistically associated with the word 'tiger', so it seems that some idea that 'electricity' (i.e. electric charge or charges) is capable of flow or motion *is* linguistically associated with

'electricity'. And perhaps this is all – apart from being a physical magnitude or quantity in the sense described before – that is linguistically associated with the word.

Now then, if anyone knows that 'electricity' is the name of a physical quantity, and his use of the word is connected by the sort of causal chain I described before to an introducing event in which the causal description given was, in fact, a causal description of electricity, then we have a clear basis for saying that he uses the word to refer to electricity. Even if the causal description failed to describe electricity, if there is good reason to treat it as a mis-description *of electricity* (rather than as a description of nothing at all) – for example, if electricity was described as the physical magnitude with such-and-such properties which is responsible for such-and-such effects, where in fact electricity is responsible for the effects in question, and the speaker intended to refer to the magnitude responsible for those effects, but mistakenly added the incorrect information 'electricity has such-and-such properties' because he mistakenly thought that the magnitude responsible for those effects had those further properties – we still have a basis for saying that both the original speaker and the persons to whom he teaches the word use the word to refer to electricity.

If a number of speakers use the word 'electricity' to refer to electricity, and, in addition, they have the standard sorts of associations with the word – that it refers to a magnitude which can move or flow – then, I suggest, the question of whether it has 'the same meaning' in their various idiolects simply does not arise. If a word is linguistically associated with a necessary and sufficient condition in the way that 'bachelor' is, then that sort of question *can* arise; but it does not arise, for example, in the case of proper names, and it does not arise, for a similar reason, in the case of physical magnitude terms. Thus if you know that 'Quine' is a name and I know that 'Quine' is a name and, in addition, we both refer to the same person when we use the word (even if the causal chains linking us to the referent are quite different) then the question of whether 'Quine' has the same meaning in my idiolect and in yours does not arise. More precisely: if the referent is the same, and we both associate the same minimal linguistic information with the word 'Quine', namely that it is a person's name, then the word is treated as the same word whether it occurs in your idiolect or in mine. Similarly, 'electricity' is the same word in Ben Franklin's idiolect and in mine. Of course, if you had wrong linguistic ideas about the name 'Quine' – for example, if you thought 'Quine' was a female name (not just that Quine was a woman, but that the name was restricted to females) – then there would be a difference in meaning.

This account stresses causal descriptions because physical magnitudes are invariably discovered through their effects, and so the natural way to first single out a physical magnitude is as the magnitude responsible for certain effects. Of course, the words 'responsible', 'causes', etc., do not literally have to occur in the description: *spin*, for example, was introduced by describing it as a physical magnitude having half-integral values characteristic of certain elementary particles, and giving a *law* connecting it with magnitudes previously introduced; I intend the notion of a causal description to include this case. And it is not a 'necessary truth' that the description introducing a new physical magnitude should involve a notion of cause or law; but I am not trying in this paper to state 'necessary truths'.

Once the term 'electricity' has been introduced into someone's vocabulary (or into his 'idiolect', as the dialect of a single speaker is called) whether by an introducing event, or by his learning the word from someone who learned it via an introducing event, or by his learning the word from someone linked by a chain of such transmissions to an introducing event, the referent in that person's idiolect is also fixed, even if no knowledge that that person has fixed it. And once the referent is fixed, one can use the word to formulate any number of theories about that referent (and even to formulate theoretical definitions of that referent which may be correct or incorrect scientific characterizations of that referent), without the word's being in any sense a different word in different theories. Thus the account just given fullfils the desideratum with which we started – it makes such terms as 'electricity' trans-theoretical. The 'operational criteria' you can give for the presence of electricity will depend strongly on what theory you accept; but, without the illicit identification of meaning with operational criteria, it does not follow at all that *meaning* depends on the theory you accept.

The possibility of formulating definite descriptions (or even misdescriptions) of physical magnitudes depends upon the availability in our language of such 'broad spectrum' notions as *physical magnitude* and *causes*; that these play a crucial role in the introduction of physical magnitude terms was argued in chapter 13, volume 1. In that paper, however, I did not distinguish between *defining* what I then called theoretical terms and *introducing* them. Of course, if we have available a language in which we can formulate descriptions of the referents of our various physical magnitude terms, then we can consider the various theories that we have containing those terms as so many different systems of sentences in that one language. To the extent that we can do this, we can treat the notions of reference and truth appropriate to that language as trans-theoretical notions also.

B. Kripke's theory of proper names

I have already acknowledged a heavy indebtedness to Kripke's (unpublished) work on proper names. Since I have heard mainly secondhand reports of that work, I shall not attempt to describe it here in any great detail. But, as it has come down to me, the key idea is that a person may use a proper name to refer to a thing or person X even though he has *no* true beliefs about X. For example, suppose someone asks me who Quine is, and I falsely tell him that Quine was a Roman emperor. If he believes me, and if he goes on to use the word 'Quine' with the intention of referring to the person to whom *I* refer as Quine, then he will say such things as 'Quine was a Roman emperor' – and he will be referring to a contemporary logician. Of course, he still has some true beliefs about Quine (beyond the belief that Quine is or was a person); for example, that Quine is or was named 'Quine'; but Kripke has more elaborate examples to show that even this is not always the case. On Kripke's view, the essential thing is this: that the use of a proper name to refer involves the existence of a causal chain of a certain kind connecting the user of the name (and the particular event of his using the name) to the bearer of the name.

Now then, I do not feel that one should be quite as liberal as Kripke is with respect to the causal chains one allows. I do not see much point, for example, in saying that someone is referring to Quine when he uses the name 'Quine' if he thinks that 'Quine' was a Roman emperor, and that is all he 'knows' about Quine; unless one has *some* beliefs about the bearer of the name which are true or approximately true, then it is at best idle to consider that the name refers to that bearer in one's idiolect. But what seems right about Kripke's account is that the knowledge an individual user of a language has need not at all fix the reference of the proper names in that individual's idiolect; the reference is fixed by the fact that that individual is causally linked to other individuals who were in a position to pick out the bearer of the name, or of some names from which the name descended. Indeed, what is important about Kripke's theory is not that the use of proper names is 'causal' – what is not?– but that the use of proper names is *collective*. Anyone who uses a proper name to refer is, in a sense, a member of a collective which had 'contact' with the bearer of the name: if it is surprising that a particular member of the collective need not have had such contact, and need not even have any good idea of the bearer of the name, it is only surprising because we think of language as private property.

The relationship of this theory of Kripke's to the above theory of physical magnitude terms should be obvious. Indeed, one might say that

physical magnitude terms *are* proper names: they are proper names of *magnitudes* not *things* – however, this would be wrong, I think, since some physical magnitude terms (e.g. 'heat') are linguistically associated with rather rich information about the referent. The important thing about proper names is that it would be ridiculous to think that having linguistic competence can be equated in their case with knowledge of a necessary and sufficient condition – thus one is led to search for something other than the knowledge of the speaker which fixes the referent in their case.

It will be noted that I required a causal chain from the use of the physical magnitude term back to an introducing event – not back to an event in which the physical magnitude played a significant role. The reason is that, although no one in practice is going to be in a position to give a definite description of a physical magnitude unless he is causally connected to such an event, the nature of *that* causal chain seems not to matter. As long as one is in a position to give a definite description (or even a misdescription), one is in a position to introduce the term; and the chain from there on is something about which much more definite statements can be made. (In my opinion, it would be good to make a similar modification in Kripke's theory of proper names.)

C. Natural kind words

In chapter 8 of this volume I presented an account of natural kind words (e.g. 'lemon') which has some relation to the present account of physical magnitude terms. I suggested that anyone who has linguistic competence in connection with 'lemon' satisfies three conditions: (1) He has implicit knowledge of such facts as the fact that 'lemon' is a concrete noun, that it is the 'name of a fruit', etc. – information given by classifying the word under certain natural syntactic and semantic 'markers'. I criticized Jerrold Katz for the view that natural systems of semantic markers can enable us to give the exact meaning of each term (or of *any* natural kind term); but *some* of the information associated with a word can naturally be represented by classifying the word under such familiar headings as 'noun', 'concrete', etc. (2) He associates the word with a certain 'stereotype' – yellow color, tart taste, thick peel, etc. (3) He uses the word to *refer* to a certain natural kind – say, a natural kind of fruit whose most essential feature, from a biologist's point of view, might be a certain kind of DNA.

Two points were most important in the argument of that paper. The first was that the properties mentioned in the stereotype (and, I would add, the properties indicated by the semantic markers) are not being

analytically predicated of each member of the extension, or, indeed, of any members of the extension. It is not analytic that all tigers have stripes, nor that some tigers have stripes; it is not analytic that all lemons are yellow, nor that some lemons are yellow; it is not even analytic that tigers are animals or that lemons are fruits. The stereotype is *associated* with the word; it is not a necessary and sufficient condition for membership in the corresponding class, nor even for being a normal member of the corresponding class. Engels' example of the word 'fish' fits right in here: what Engels was pointing out was precisely that the stereotype associated with the term 'fish' even in scientific, as opposed to lay, usage is not a necessary and sufficient condition. The second point was that speakers must be referring to a particular natural kind for us to treat them as using the same word 'lemon', or 'aluminum', or whatever. The weakness of that paper, apart from being very poorly organized and presented, is that nothing positive is said about the conditions under which a speaker who uses a word (say 'aluminum' or 'elm tree') is referring to one set of things rather than another. Clearly, the speaker who uses the word 'aluminum' need not be able to tell aluminum from molybdenum, and the speaker who use the term 'elm tree' cannot tell elm trees from beech trees if he happens to be me. But then what does determine the reference of the terms 'aluminum', and 'molybdenum' in my idiolect? In the previous papers, I suggested that the reference is fixed by a test known to experts; it now seems to me that this is just a special case of my use being causally connected to an introducing event. For natural kind words too, then, linguistic competence is a matter of knowledge plus causal connection to introducing events (and ultimately to members of the natural kind itself). And this is so far the same reason as in the case of physical magnitude terms; namely, that the use of a natural kind word involves in many cases membership in a 'collective' which has contact with the natural kind, which knows of tests for membership in the natural kind, etc., only as a collective. The idea that linguistic competence in connection with a natural kind word involves more than just having the right extension or reference (where this is now explained via a causal account), but also associating the right stereotype seems to me to carry over to physical magnitude words. Natural kind words can be associated with 'strong' stereotypes (stereotypes that give a strong picture of a stereotypical member – even to the point of enabling one to tell, in most cases, if something belongs to the natural kind), as in the case of 'lemon' or 'tiger', or with 'weak' stereotypes (stereotypes that give no idea of what a sufficient condition for membership in the class would be), as in the case of 'molybdenum' or (unless I am a very atypical speaker) 'elm'. Similarly, it seems to me that the physical

magnitude term 'temperature' is associated with a very strong stereotype, and 'electricity' with a weak one.

D. Objections and questions

It is obvious that the account presented here must face certain hard questions. Without attempting to think of all of them myself, I should like to list a few that may help to launch discussion.

(1) One question that must be faced by all causal theories of meaning is how to make more precise the notion of a causal chain of the appropriate kind. How precisely can we describe the sorts of causal chains that must exist from one use of a word to a later use of the same word if we are to say that the referent or referents are the same in the two cases? And how much of a defect in these sorts of theories is it if one cannot be more precise on this point?

(2) It may seem counterintuitive that a natural kind word such as 'horse' is sharply distinguished from a term for a fictitious or nonexistent natural kind such as 'unicorn', and that a physical magnitude term such as 'electricity' is sharply distinguished from a term for a fictitious or nonexistent physical magnitude or substance such as 'phlogiston'. Indeed, I myself believe that if unicorns were found to exist and people began to discover facts about them, give nonobvious definite descriptions or approximately correct descriptions of the class of unicorns, etc., then the linguistic character of the word 'unicorn' would change; and similarly with 'phlogiston'; but this is certain to be controversial.

(3) Some people will argue that definitions of such terms as 'electricity' (or, more precisely, 'charge') are crucial in the exact sciences, and further that such definitions should be regarded as *meaning stipulations*. I agree with the first part of this – that definitions are important in science, provided one remembers what Quine has pointed out, that 'definition' is relative to a particular text or presentation, and that there is no such thing, in general, as the definition of a term 'in physics' or 'in biology' – only the definition in *X*, *Y*, or *Z*'s presentation or axiomatization. I disagree with the last part – that 'definitions' in science are meaning stipulations – but, again, this is certain to be controversial.

(4) Finally, there will be objections to my use of causal notions, from Humeans who expect them to be reduced away, and to my use of the term 'physical magnitude' from extensionalists and nominalists. Here I can only plead guilty to the belief that talk about what causes what, or what the laws of nature are, or what would happen if other things happened is *not* highly derived talk about mere regularities, and to the

belief that the real world requires for its description not only reference to things but reference to physical magnitudes (cf. chapter 19, volume 1 of these papers) – in a sense of 'physical magnitude' in which physical magnitudes exist contingently, not as a matter of logical necessity, and in which magnitudes can be synthetically identical (e.g. temperature is the same magnitude as mean molecular kinetic energy).

III. Why positivistic theory of science is wrong

My contention in this paper is not that what is wrong with positivist theory of science is positivist theory of meaning. What is wrong with positivist theory of science is that it is based on an idealist or idealist-tending world view, and that that view does not correspond to reality. However, the idealist element in contemporary positivism enters precisely through the theory of meaning; thus part of any realist critique of positivism has to include at least a sketch of rival theory. In the present section, I want to turn from the task of sketching such a rival theory, which was just completed, to the task of showing that positivistic theory of explanation broadly construed – that is, positivist theory of scientific theory – does not correspond to reality any better than the older and less sophisticated idealist theories to which it is historically the successor.

Let us for a moment review some of those older theories. The oldest theory is Bishop Berkeley's. Here one already meets what might be called the *adequacy claim*: that is, the claim that a convinced Berkelian is *entitled* to accept standard scientific theory and practice, that Berkeley can give an account of the scientific method which would justify this. Indeed, I have heard philosophers argue that acceptance of Berkeley's metaphysics would not make any difference to the scientific theories one would accept. Here one already meets an important ambiguity. One can be claiming that a Berkelian can make the move of 'accepting' scientific theory in some sense other than accepting it as true or approximately-true: say, accepting it as a useful prediction heuristic. If this is what one means, then the claim is trivial. To be sure, Berkeley can 'accept' Newtonian physics in the Pickwickian sense of 'accept' as a useful scheme for making predictions. But Berkeley, to do him justice, was interested in much more: what he claimed was that an idealist could *reinterpret* (only he would not consider it *re*-interpretation, but rather *correct* interpretation) the notion of object so as to square both the layman's and the scientist's talk of objects with the idealist claim that reality consists of minds and their sensations ('spirits' and their 'ideas').

The difference between the two claims is the difference between accepting the idea that social practice is the test of truth and rejecting it, between accepting the idea that the overwhelming success of scientific theory offers some reason for accepting that theory as true or approximately-true, and claiming that success in practice is *no* indication of truth. Machian positivism fails for the same reason that Berkelian idealism does: although Mach makes the claim that his construction of the world out of sensations ('Empfindungen') is compatible with lay and scientific object-talk, no demonstration at all is given that this is so. The first philosopher to both precisely state and to undertake the task of *translating* thing-language into phenomenalistic language was Carnap (in *Logische Aufbau der Welt*). And what does Carnap do? He devotes the entire book to *preliminaries*, to 'reconstructions' *within* sensationalistic language (i.e. reductions of some sensation-concepts to others, not of thing-concepts to sensation-concepts), and then in the last chapter gives a sketch of the relation of thing-language to sensation-language which is *not* a translation, and which, indeed, amounts to no more than the old claim that we pick the thing-theory that is 'simplest' and most useful. In short, no demonstration is given at all that the positivist is entitled to quantify over (or refer to) material things.

It is with the failure of the phenomenalist translation enterprise, that is, with the failure to find *any* interpretation of object-concepts under which the prima facie incompatibility between an idealist world-view and a materialist world-view, between a world consisting of 'spirits and their ideas', or of 'Empfindungen', or of total experience-slices in one 'specious present', and a world consisting of fields and particles, simply *disappears* – it is with this failure that contemporary positivistic philosophy of science begins. Basically, two moves were made by the positivists after the failure of phenomenalist translation. The first was to give up construing scientific theories as systems of statements each of which had to have an intelligible interpretation (intelligible from the standpoint of what was taken as 'completely understood' or 'fully interpreted'), and to construe them rather as mere calculi, whose objective was to give successful predictions and otherwise to be as 'simple' as possible. 'Scientific theories are partially interpreted calculi' (chapter 13, volume 1 of these papers). The second move was to shift from phenomenalist language to 'observable thing language' as one's reduction-base – i.e. to say that one was seeking an interpretation or 'partial interpretation' of physical theory in 'observable thing language', not in 'sensationalistic language'.

The second move may make it appear questionable whether positivism is still correctly characterized as an 'idealist' tendency – i.e. as a tendency which regards or tends to regard the 'hard facts' as just facts about actual and potential *experiences*, and all other talk as somehow just highly derived talk about actual and potential experiences. I, myself, think this characterization *is* still fundamentally correct despite the shift to 'observable thing predicates' for two reasons: (1) The cut between observable things and 'theoretical entities' was historically introduced as a substitute for the thing/sensation dichotomy. Indeed, the reduction of 'theoretical entities' to 'observable things and qualities' would hardly seem to be a natural problem to someone who did not have in the back of his head the older problem of reduction to *sensations*. The reduction of things to sensations is both a historically motivated problem and one which rests upon the sharpness of the distinction between a material thing and a sensation (of course, even this sharpness is partly an illusion, in a materialist view – substitute 'material process' for 'material thing'!), as well as the supposed 'certainty' one has concerning one's own sensations. But the reduction of electrons to tables and chairs, or, more generally, of 'unobservable' things to 'observable' things is not historically motivated, the distinction is not sharp (Grover Maxwell asked years ago if a dust note is something 'given' when it is just big enough to see and a 'construct' when it is just too small to see – can the distinction between data and construct be a matter of size?), and one is not supposed to have certainty concerning observable things. (2) The positivists themselves frequently say that one could carry their analysis back down to the level of sensations, and that stopping with 'observable thing predicates' is a matter of *convenience*.†

In the remainder of this section I want to show that the first move – construing scientific theories as partially interpreted calculi – does not solve the adequacy problem at all. The positivist today is no more entitled than Berkeley was to accept scientific theory and practice – that is, his own story leads to no reason to think either that scientific theory is true, or that scientific practice tends to discover truth. In a sense, this is immediate. The positivist does not claim that scientific theory is 'true' in any trans-theoretic sense of 'true'; the only trans-theoretic notions he has are of the order of 'leads to successful prediction' and 'is simple'. Like the Berkelian, he has to fall back on the position that scientific theory is *useful* rather than true or approximately-true. But he does try to provide some account of the acceptability of scientific theories, even some account of their 'interpretation'. And he wants to maintain that in some sense the principle on which realist philosophy of science rests – that social practice is the test of truth, that the success of scientific theories is reason to think they are true or approximately-true –

† E.g. Carnap says this on p. 63 in Carnap (1956).

is right. What I want to show is that the notion of 'truth' that the positivist can give us is not the one on which scientific practice is based.

A. Truth

When a realistically minded scientist – that is to say, a scientist *whose practice* is realistic, not one whose official 'philosophy of science' is realistic – accepts a theory, he accepts it as true (or probably true, or approximately-true, or probably approximately-true). Since he also accepts *logic*† he knows that certain moves *preserve truth*. For example, if he accepts a theory T_1 as true and he accepts a theory T_2 as true, then he knows that T_1 & T_2 – the *conjunction* of T_1 and T_2 – is also true, by logic, and so he accepts T_1 & T_2. If we talk about probability, we have to say that if T_1 is very highly probably true and T_2 is very highly probably true, then the conjunction T_1 & T_2 is also highly probable (though not *as* highly as the conjuncts separately), provided that T_1 is not negatively relevant to T_2 – i.e. provided that T_2 is not only highly probable on the evidence, but also no less probable on the added assumption of T_1 (this is a judgement that must be made on the basis of what T_1 *says* and of background knowledge, of course). If we talk about approximate-truth, then we have to say that the approximations probably involved in T_1 and T_2 need to be compatible for us to pass from the approximate-truth of T_1 and T_2 to the approximate-truth of their conjunction. None of these matters is at all deep, from a realist point of view. But even if we confine ourselves to the simplest case, the case in which we can neglect the chances of error and the presence of approximations, and treat the acceptance of T_1 and T_2 as simply the acceptance of them as true, I want to suggest that the move from this acceptance to the acceptance of the conjunction is one to which one is not entitled on positivist philosophy of science. One of the simplest moves that scientists daily make, a move they make as a matter of propositional logic, a move which is central if scientific inquiry is to have any *cumulative* character at all, is totally *arbitrary* if positivist philosophy of science is right.

The difficulty is very simple. Acceptance of T_1, for a positivist, means acceptance of the calculus T_1 as leading to successful predictions (i.e. all *observation sentences* which are theorems of T_1 are true; not all *sentences* which are theorems of T_1 are 'true' in any fixed trans-theoretic sense). Similarly, the acceptance of T_2 means the acceptance of T_2 as leading to successful predictions. But from the fact that T_1 leads to successful predictions and the fact that T_2 leads to successful predictions it does not follow at all that the conjunction T_1 & T_2 leads to successful predictions. The difficulty, in a nutshell, is that the predicate which plays the *role* of truth – the predicate 'leads to successful predictions' – does not have the *properties* of truth. The positivist may teach in his philosophy seminar that acceptance of a scientific theory is acceptance of it as 'simple and leading to true predictions', and then go out and do science (or his students may go out and do science) by verifying theories T_1 and T_2, conjoining theories which have been previously verified, etc. – but then there is just as great a discrepancy between what he teaches in his philosophy seminar and his *practice* as there was between Berkeley's teaching that the world consisted of spirits and their ideas and continuing in practice to daily rely on the material object conceptual system.

† The role of logic in empirical science is discussed in Putnam (1971) and in chapter 10, volume 1 of these papers.

Nor does it help to bring in 'simplicity'. It is not obvious that the conjunction of simple theories is simple; and even if simplicity is preserved, the conjunction of simple theories which separately lead to no false predictions may even be *inconsistent* (examples are easy to construct). More sophisticated moves have indeed been made. Thus, for Carnap truth of a theory is the same as truth of its 'Ramsey sentence' (for details see Hempel, 1965). But exactly the same objection applies: 'truth of the Ramsey sentence' does not have the properties of truth: if T_1 has a true Ramsey sentence and T_2 has a true Ramsey sentence it does not at all follow that the conjunction does.

(For those readers familiar with Carnap's use of the Hilbert epsilon-symbol, it may be pointed out that the difficulty comes out in very sharp form in Carnap's symbolization of his interpretations of individual theoretical terms. Thus let $T_1(P)$, $T_2(P)$ be two theories containing exactly one theoretical term P. On Carnap's own symbolization of his view, what P means in T_1 is $\varepsilon P T_1(P)$; what P means in T_2 is $\varepsilon P T_2(P)$; and what P means in T_1 & T_2 is $\varepsilon P[T_1(P)$ & $T_2(P)]$; this makes it explicit that P has different meanings in T_1 and T_2 and *yet a third meaning* in their conjunction.)

B. Simplicity

It is easy to construct a 'theory' in the positivist sense (a calculus containing some observation terms) which leads to no false predictions but which no scientist would dream of accepting. This is usually handled by saying that scientists only choose 'simple' theories. Also, a simple theory may mess up science as a whole: so it is said that scientists are trying to maximize the simplicity of 'total science'. 'Theory' means, then, 'formalization of total science, or of some piece which is independent of the rest of total science'. Unfortunately, no one has ever written down or ever will write down a 'theory' in this sense. The fact is, that positivist

philosophy of science depends on a constant slide between giving the impression that one is talking about 'theories' in the customary sense – Newton's theory, Maxwell's theory, Darwin's theory, Mendel's theory – and saying, at key points of difficulty such as the one just alluded to, that one is *really* talking about a 'formalization of total science', or some such thing.

The difficulty with the rule 'choose the simplest theory compatible with the evidence' is that it is probably not *right*, or would probably not be right, even if one *could* formalize 'total science' (at a given time). Scientists are not trying to maximize some formal property of 'simplicity'; they are trying to maximize *truth* (or improve their approximation to truth, or increase the amount of approximate-truth they know without decreasing the goodness of the approximation, and so forth).

Of course, a realist might accept the rule 'choose the simplest hypothesis', if it could be shown that the simplest hypothesis is always the most *probable* on the basis of the rest of his knowledge. But this is not so on any usual measure of simplicity. For example, suppose I know just three points on interstate highway 40, and those three points lie on a straight line. Suppose also that the statement 'IS 40 is straight' is logically consistent with my total knowledge. Then accepting 'IS 40 is straight' would, on the usual simplicity metrics, be accepting the simplest hypothesis. Yet I would not in fact accept 'IS 40 is straight', nor would anyone with our background knowledge. Given that every other interstate highway has curves, and given the enormous length of IS 40 and the enormous impracticality of making a straight highway across the entire United States, it is overwhelmingly probable that IS 40 is *not* straight.

Can we not say that my *total* 'knowledge' is less simple if I accept 'IS 40 is straight'? Not, it seems to me, on the basis of any criterion of *simplicity* that I know of. What is obviously involved here is not *simplicity* but plausibility: what introducing the word 'simplicity' does is make it look as if a calculation which is in fact the calculation of the probability of a state of affairs is in reality just a calculation of a formal property (such as number of argument places, number of primitive symbols, length and number of the axioms, perhaps shape of the curves mentioned) of an uninterpreted or semi-interpreted *calculus*, even if the property of being the most probable hypothesis on background knowledge could be *represented* syntactically, omitting to mention that the representing property was the syntactic representation of a *probability measure*, and pretending that it was *just* a formal property (like having simple axioms), would be a way of disguising rather than revealing what was going on.

C. Confirmation

Indeed, positivist philosophers of science have made attempts at formalizing the logic of confirmation. These attempts are interesting (though so far unsuccessful) researches on *any* philosophy of science. But not only do they have nothing to do with positivist theory of meaning; they are in fact *incompatible* with it. Thus when they write about meaning, positivists tell us that 'theoretical terms' have different meanings in different theories; when they formalize confirmation theory, they invariably treat theories as systems of sentences in *one* language, and assume that all semantical concepts are *trans*-theoretic. Thus the positivists are engaged in formalizing *realistic* confirmation theory: not the confirmation theory (if there is one!) to which their own theory of meaning should lead.

What is going on here should be evident from Carnap's work on the foundations of mathematics. Carnap has a consistent tendency to *identify* concepts with their syntactic representations: thus, mathematical truth with theoremhood (after the discovery of Gödel's theorem, he either allowed 'nonconstructive rules of proof', or simply assumed set theory, and took 'logical consequence' rather than derivability as the basic notion, although this trivialized the 'analysis' of mathematical truth). In the same way he would have liked to identify a state of affairs having a probability of, say, 0.9, with the corresponding sentence's having a c-value of 0.9 (where 'c' would be a syntactically defined measure on sentences in a formalized language). Even if Carnap had found a successful 'c-function', the fact is that it would have been successful because it corresponded to a reasonable probability measure over some collection of states of affairs; but this is just what Carnap's positivism did not allow him to say.

D. Auxiliary hypotheses

Sometimes, as we mentioned, the positivists make it explicit that the 'theories' to which their theory of science applies are 'formalizations of total science', and not theories in the usual sense; but their readers do, I think, tend to come away with the impression that their model *is* a model of a scientific theory in the usual sense – especially, a physical theory. Believing this involves believing that a physical theory is a calculus, or could easily be formalized as a calculus, and that its predictions are *self-contained* – that they are deduced from the explicitly stated assumptions of the theory itself. This leads to a comparison with social sciences which is derogatory to the social sciences – for the classic social science theories are clearly *not* self-contained in this sense. In

short, the positivist attitude tends to be that social science is science only when and to the extent that it apes *physics*. And this for the reason that the mathematical model of a scientific theory provided by the positivists is thought to clearly fit *physical* theories.

But, in fact, it fits physical theories very badly, and this for the reason that even physical theories in the usual sense – e.g. Newton's Theory of Universal Gravitation, Maxwell's theory – lead to no predictions at all without a host of auxiliary assumptions, and moreover without auxiliary assumptions that are not at all law-like, but that are, in fact, assumptions about boundary conditions and initial conditions in the case of particular systems. Thus, if the claim that the term 'gravitation', for example, had a meaning which depended on the theory were true, and the theory included such auxiliary assumptions as that 'space is a hard vacuum', and 'there is no tenth planet in the solar system', then it would follow that discovery that space is *not* a hard vacuum or even that there is a tenth planet would change the meaning of 'gravitation'. I think one has to be pretty idealistic in one's intuitions to find this at all plausible! It is not so implausible that knowledge of the meaning of the term 'gravitation' involves some knowledge of the theory (although I think that this is wrong: the stereotype associated with 'gravitation' is not nearly as strong as a particular theory of gravitation), and this is probably what most readers think of when they encounter the claim that physical magnitude terms (usually called 'theoretical terms' to prejudge just the issue this paper discusses) are 'theory loaded'; but the actual meaning-dependence required by positivist meaning theory would be a dependence not just on the *laws* of the theory, but on the particular auxiliary assumptions – for, if these are not counted as part of the theory, then the whole theory-prediction scheme collapses at the outset.

Finally, neglect of the role that auxiliary assumptions actually play in science leads to a wholly incorrect idea of how a scientific theory is confirmed. Newton's theory of gravitation was not confirmed by checking predictions derived from it plus some set of auxiliary statements fixed in advance; rather the auxiliary assumptions had to be continually modified and expanded in the history of Celestial Mechanics. That scientific problems as often have the form of finding auxiliary hypotheses as they do of finding and checking predictions is something that has been too much neglected in philosophy of science;† this neglect is largely the result of the acceptance of the positivist model and its uncritical application to actual physical theories.

† I discuss this in chapter 16, volume 1 of these papers.

12

The meaning of 'meaning'*

Language is the first broad area of human cognitive capacity for which we are beginning to obtain a description which is not exaggeratedly oversimplified. Thanks to the work of contemporary transformational linguists,† a very subtle description of at least some human languages is in the process of being constructed. Some features of these languages appear to be *universal*. Where such features turn out to be 'species-specific' – 'not explicable on some general grounds of functional utility or simplicity that would apply to arbitrary systems that serve the functions of language' – they may shed some light on the structure of mind. While it is extremely difficult to say to what extent the structure so illuminated will turn out to be a universal structure of *language*, as opposed to a universal structure of innate general learning strategies,‡ the very fact that this discussion can take place is testimony to the richness and generality of the descriptive material that linguists are beginning to provide, and also testimony to the depth of the analysis, insofar as the features that appear to be candidates for 'species-specific' features of language are in no sense surface or phenomenological features of language, but lie at the level of deep structure.

The most serious drawback to all of this analysis, as far as a philosopher is concerned, is that it does not concern the meaning of words. Analysis of the deep structure of linguistic forms gives us an incomparably more powerful description of the *syntax* of natural languages than we have ever had before. But the dimension of language associated with the word 'meaning' is, in spite of the usual spate of heroic if misguided attempts, as much in the dark as it ever was.

In this essay, I want to explore why this should be so. In my opinion, the reason that so-called semantics is in so much worse condition than syntactic theory is that the *prescientific* concept on which semantics is

* First published in K. Gunderson (ed.) *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science, VII (University of Minnesota Press, Mpls.)

† The contributors to this area are now too numerous to be listed: the pioneers were, of course, Zellig Harris and Noam Chomsky.

‡ For a discussion of this question see Putnam (1967) and N. Chomsky (1971), especially chapter 1.