Mind, Language and Reality

Philosophical Papers, Volume 2

HILARY PUTNAM

Professor of Philosophy, Harvard University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1975
First paperback edition 1979
Reprinted 1980, 1982, 1984, 1986, 1987, 1989, 1992, 1993,
1995. 1997

Typeset in Imprint

Library of Congess Catalog card number: 75-8316

ISBN 0-521-29551-3 paperback

Transferred to digital printing 2003

Contents

Volume 2

Introduction: Philosophy of language and the rest of philosophy

For over a hundred years, one of the dominant tendencies in the philosophy of science has been verificationism: that is, the doctrine that to know the meaning of a scientific proposition (or of any proposition, according to most verificationists) is to know what would be evidence for that proposition. Historically, verificationism has been closely connected with positivism: that is, at least originally, the view that all that science really does is to describe regularities in human experience. Taken together, these views seem close to idealism. However, many twentieth-century verificationists have wanted to replace the reference to experience in the older formulations of these doctrines with a reference to 'observable things' and 'observable properties'. According to this more recent view, scientific statements about the color of flowers or the eating habits of bears are to be taken at face value as referring to flowers and bears; but scientific statements about such 'unobservables' as electrons are not to be taken as referring to electrons, but rather as referring to meter readings and the observable results of cloud chamber experiments. It is not surprising that philosophers who took this tack found themselves in a certain degree of sympathy with psychological behaviorism. Just as they wanted to 'reduce' statements about such unobservables as electrons to statements about 'public observables' such as meter readings, so they wanted to reduce statements about phenomena which, whatever their private status, were publicly unobservable, such as a person's sensations or emotions, to statements about such public observables as bodily behaviors.

At this point, they found themselves in a certain bind. On the one hand, the doctrine that talk about sensations or emotions is simply talk about a person's behavior is so implausible that almost no philosopher has been able to maintain it, or at least to maintain it for long. On the other hand, if the intuition behind recent verificationism is right, and to know the meaning of a statement is to know what would be *public* evidence for it, then it seems as if there has to be something right about behaviorism. And so philosophers tried to develop a philosophy to this effect – a philosophy that would say that 'naive behaviorism' was false but that nevertheless there was *some* kind of semantical or logical

relation between statements about emotions and feelings and statements about behavior.

In my opinion, verificationism and behaviorism are fundamentally misguided doctrines. In the first volume of these collected papers I have tried to do a certain amount of philosophy of science from a nonverificationist and nonpositivist point of view, but without developing in detail a theory of meaning alternative to the positivists'. The papers in the present volume, while written over a number of years and betraying a number of changes of mind, have been largely concerned with the development of such a theory of meaning, a nonverificationist theory of meaning, and with the critique of verificationist philosophy of mind.

The defects of verificationism

One of the defects of verificationism that was early noticed by the more sophisticated verificationists themselves, and especially by Hans Reichenbach, was a certain distortion of the character of actual scientific methodology and inference. Naive verificationism would say that the statement 'There is current flowing in this wire' means 'The voltmeter needle is displaced', or something of that kind. That is, the relation between the so-called theoretical statement that current is flowing in the wire and the evidence for it is assimilated to the relation between 'John is a bachelor' and 'John is a man who has never been married'. Now the latter relation is itself not as simple a thing as it may seem at first blush (cf. the paper 'The analytic and the synthetic' in this volume), but it is roughly right that the relation is a *conventional*† one: 'John is a bachelor' is equated by some kind of conventional agreement with 'John is a man who has never been married'. But, as Reichenbach pointed out in *Experience and Prediction*, the relation between the theoretical statement and the evidence for it (say, 'There is current flowing in the wire' and 'The voltmeter needle is displaced') is a probabilistic inference within a theory. It is not that we equate the *sound-sequence* 'There is current flowing in this wire' with 'The voltmeter needle is displaced' by an act of conventional stipulation; it is rather that we accept a theory of electricity and of the structure of voltmeters from which it follows that, with a high probability, the voltmeter needle will be displaced if there is current flowing in the wire, and vice versa. To represent what are in fact probabilistic inferences within theories as logical equivalences is a serious distortion. To

† The conventionality of analytical sentences is well explicated, in my opinion, in Lewis (1969).

represent these inferences as purely conventional meaning equivalences is an even more serious distortion.

Some of the criticisms that I make of behaviorism in this volume really require little more than the critique of naive verificationism just alluded to. In particular, 'Dreaming and "depth grammar"', and 'Brains and behavior' represent criticisms of philosophical behaviorism from a nonverificationist standpoint; but those criticisms would be accepted, I believe, by a sophisticated verificationist like Reichenbach or Carnap.

But sophisticated verificationism found that it had escaped from one difficulty to land in another. If *meaning* is conflated or confounded with *evidence*, and what is evidence for a statement is a function of the total theory in which the statement occurs, then every significant change in theory becomes a change in the meaning of all the constituent words and statements of the theory. One of the early verificationists, Charles Peirce, anticipated this difficulty in the last century when he came to the conclusion that every change in a person's 'information' is a change in 'the meaning of his words'. But the distinction between the meaning of a man's words and what he believes about the facts, the distinction between disagreement in the meanings of words and disagreement about the facts, is precisely central to any concept of linguistic *meaning*. If we come to the conclusion that that distinction is untenable then, as Quine has long urged, we should abandon the notion of meaning altogether. With the exception of Quine, most verificationists have found this course unattractive. Thus they were caught in a serious dilemma – caught between their desire to continue talking about meaning in something like the traditional way, and their adherence to the network theory of meaning which taken seriously implies that nothing can be made of the notion of linguistic meaning.

For a realist, the situation is quite different. No matter how much our theory of electrical charge may change, there is one element in the meaning of the term 'electrical charge' that has not changed in the last two hundred years, according to a realist, and that is the reference. 'Electrical charge' *refers to the same magnitude* even if our theory of that magnitude has changed drastically. And we can identify that magnitude in a way that is independent of all but the most violent theory change by, for example, singling it out as the magnitude which is causally responsible for certain effects.

But the realist has his problems too. Traditionally realists thought that reference was determined by mental or Platonic entities, intensions. This doctrine of fixed 'meanings', either in the head or in the realm of abstract entities (and somehow connected to the head), determining reference once and for all, is open, interestingly enough, to some of the

same objections that can be brought to bear against verificationism.

Thus, very recently realists have begun to redevelop their theory of meaning. Instead of seeing meanings as entities which determine reference, they now are beginning to see meanings as largely determined by reference, and reference as largely determined by causal connections. This sort of nonverificationist theory of meaning is presented briefly in 'Explanation and reference' and at more length in 'The meaning of "meaning"'.

I would not wish to give the impression that the only problem with verificationism is its inability to give a correct account of our customary notion of meaning, however. Truth and falsity are the most fundamental terms of rational criticism, and any adequate philosophy must give some account of these, or failing that, show that they can be dispensed with. In my opinion, verificationism has not succeeded in doing either. There is a sense in which Tarski's technical work in mathematical logic enables one to explicate the notion of truth in the context of a language with fixed meanings, and as long as there is no doubt that the terms of that language have clear reference. (Even in that context, one may question whether we have been given an account of what 'true' means, or simply a substitute for the word 'true' designed for that specific context.) But if the meaning of words is a function of the theory in which they occur, and changes as that theory changes, then if we limit ourselves to Tarski's methods, 'true' and 'false' can only be defined in the context of a particular theory. In particular, Tarskian semantics gives no explanation of the meanings of 'true' and 'false' when they are used to compare and criticize different theories, if meaning is really theory-dependent. But it is just the extra-theoretic notions of truth and falsity which are indispensible for rational criticism,† which is why they have always been taken as fundamental in the science of logic. In particular, a verificationist cannot explain why, if even the commonest scientific terms (e.g. 'voltage', 'density', 'pressure') have different meanings in the context of different theories, it should ever be justified to *conjoin* a proposition verified by one group of scientists and a proposition verified by a different group of scientists.‡ The simple fact that

† When I say that truth and falsity are the fundamental terms of rational criticism, I don't mean that we always are able to judge that a theory or doctrine is true or false; often we are lucky to be able to say that something is 'probably true' or 'approximately true'. But the semantics of probable truth and approximate truth presupposes the semantics of truth and falsity; these notions make no sense if truth and falsity make no sense.

‡Suppose the first group of scientists are experts in paleontology, and they confirm a sentence S_1 in the context of Ph (basic physics) and Pa (paleontology), then the whole theory to which S_1 belongs is S_1 & *Ph* & *Pa*. Suppose a second group of scientists are experts on radioactivity, and they confirm a sentence S_2 in the context of *Ph* and *Ad*

the conjunction of true statements is true becomes replaced by the mysterious fact that scientists are in the habit of conjoining statements which use words with different meanings and somehow, nevertheless, manage to get successful results. (This and related criticisms of verificationism are put forward in two papers in the present volume, 'Explanation and reference' and 'Logical positivism and the philosophy of mind'. In an insightful unpublished essay titled 'Realism and Scientific Epistemology', Richard Boyd has argued that these defects of verificationism and positivism are symptomatic of a deeper defect; that even if verificationism could give a correct *description* of the practice of scientists, it lacks any ideas which would enable one to explain or understand why scientific practice *succeeds*.)

Philosophy of mind

Let us now leave the topic of verificationism, and ask the more general question 'How much can the philosophy of language tell us about the philosophy of mind?' (This question is discussed in general terms in 'Language and reality' and 'Logical positivism and the philosophy of mind'.) Certain facts lie more or less on the surface. It is conceivable that one could produce an imitation of a tree that would fool even a careful observer – say, a tree made of plastic, or better, of some new synthetic material that looked and felt like bark. Thus it is not a logically necessary truth (and probably not even a truth) that anything that a normal observer who is paying attention cannot distinguish from a tree is a tree. But anything that a normal person who is paying attention cannot distinguish from a pain – that is, anything that he or she cannot distinguish from a real honest-to-God pain – is necessarily a pain. If the term *T* is used in such a way that anything that a normal person who is paying attention cannot tell from a member of the extension of T *counts* as a member of the extension of *T*, then let us say that the term T has the *appearance-logic*. In this terminology, what has just been said may be restated thus: the term 'pain', like many other sensation terms, has the appearance-logic.

The fact that many sensation terms have the appearance-logic accounts, of course, for the 'incorrigibility' of certain sentences containing these terms, such as it is. (Oddly enough, this simple and metaphysically neutral explanation of 'incorrigibility' – that many sensation terms have the appearance-logic – appears to have been overlooked by

(advanced physical theory). So S_2 belongs to the theory S_2 & *Ph* &*Ad*. Since these are different theories, any term common to S_1 and S_2 must have *different meanings* in the two contexts, if meaning is theory-dependent. So concluding that S_1 & S_2 is true (in the context S_1 & S_2 & *Pa* &*Ph* & *Ad*) from the fact that S_1 and S_2 separately have been 'verified' would be a fallacy.

many philosophers.) Some philosophers of a materialist stamp have suggested that 'I have a pain in my arm' means that I am in the sort of state (i.e. the sort of physical state) that normally produces certain effects, or the sort of state that normally has certain causes, or the sort of state that normally has certain causes and certain effects. This theory cannot be right, for it would make the statement that I have a pain in my arm a *hypothesis*, and a rather risky one at that. Similarly 'I have a pain' cannot mean that I am in a state which obeys certain psychological laws, nor can it mean that I have certain behavior dispositions, for the same reason. Certain philosophers have suggested that the word 'pain' does not have the same meaning in 'I have a pain' and in 'You have a pain', in order to avoid this argument. But this far-fetched move appears to be totally unnecessary. I think that we should, rather, take seriously the idea that the word 'pain' is a *name*. It is the name of a sensation; it has a very important reporting use; it names the very sensation that that reporting use reports.

Of course there are many problems about the theory of reference in connection with names. (Saul Kripke has made a very important contribution to this topic in his 'Naming and Necessity'.) More needs to be done on the nature of names, and of reference, and of names of sensations in particular. But one thing is clear: if sensation terms are names, then that is no need to regard them as synonymous with or in any way logically connected to *descriptions*, whether those be descriptions in terms of brain states, or descriptions in terms of psychological theories, or descriptions in terms of behavior dispositions.

Against this, one hears the following line of argument, which goes back to the later Wittgenstein: sensation terms, say 'pain', are terms we learn from other people. Other people tell whether or not a speaker is using one of those words correctly on the basis of his behavior. But, now, the criteria that other speakers use to tell whether or not a speaker is using a word correctly are connected with the meaning of that word if anything is. So behavioral criteria *must* be connected with the meaning of such words as 'pain'.

It would be instructive to go through all the things that are wrong with this argument, but for the moment let me point out that there is a certain equivocation on the notion of a *criterion* here. A criterion may be either an abstract criterion (e.g. is the speaker using the word 'pain' to refer to *pain*?); or it may be an operational criterion, (e.g. does the speaker behave in such and such a way when he reports 'I have a pain'?). That the abstract criterion states a necessary condition for having a normal usage of the word pain I do not doubt; but without assuming the truth of some form of verificationism, I see no way to get from that

harmless conclusion to the behaviorist conclusion that there is some logical relation between the statement that one is in pain and some particular behavior or behavior disposition. Of course, speakers must have some operational criteria or other to tell whether or not other speakers are using the language correctly. What is at issue is whether every change in such operational criteria has to be counted as a change in the meaning of words. In 'The meaning of "meaning"' I argue for a negative answer in the case of natural kind words; and I would similarly argue for a negative answer in the case of sensation words.

So far our conclusions are mainly negative. One cannot conclude from an examination of the meaning of psychological words that what they refer to are brain states, or that what they refer to are behavior dispositions, or that what they refer to are functional states, i.e. states characterized by psychological theories. No important theory of the nature of mind can either be confirmed or ruled out by an examination of the meanings of mental words.

Nevertheless I do argue for a particular theory of the nature of mental states in these papers. The theory for which I argue is a form of functionalism – not functionalism as a doctrine about the meanings of psychological words, but functionalism as a synthetic hypothesis about the nature of mental states.

According to functionalism, the behavior of, say, a computing machine is not explained by the physics and chemistry of the computing machine. It is explained by the machine's *program*. Of course, that program is realized in a particular physics and chemistry, and could, perhaps, be deduced from that physics and chemistry. But that does not make the program a physical or chemical property of the machine; it is an abstract property of the machine. Similarly, I believe that the psychological properties of human beings are not physical and chemical properties of human beings, although they may be realized by physical and chemical properties of human beings. Although any behavior of a computing machine that can be explained by the program of that computing machine can, in principle, be predicted on the basis of the physics and chemistry of the machine; the latter prediction may be highly unexplanatory. Understanding why the machine, say, computes the decimal expansion of π, may require reference to the abstract or functional properties of the machine, to the machine's program and not to its physical and chemical make up.

I was originally led to functionalism by a desire to defend materialism, but the considerations just mentioned seem to me to constitute a refutation of one kind of classical materialism, viz. reductionism. Although our psychological properties have their realization in our biological

make up, psychology has, if my present view is right, an *autonomous* explanatory function. This change in my view is described in the paper 'Philosophy and our mental life'. (Having come this far I was pleasantly surprised to find that my view was substantially the same as Aristotle's, although stated a bit more precisely with the aid of the vocabulary of contemporary scientific methodology and cybernetics.)

I have said that my view is a synthetic hypothesis, not a contention about the meaning of mental words. While I am reasonably convinced that my view is the correct one, I should add that I am by no means certain that it is. Indeed, it would ill behoove anyone in the present state of our knowledge to be certain of any view on so central a mystery as the relation of our bodies and souls.

The *a priori* and the analytic–synthetic distinction

In 1951, Quine caused a commotion in the community of professional philosophers by publishing an attack on the venerable distinction between analytic and synthetic propositions. In their reply to Quine, Grice and Strawson advanced two arguments: (1) when there is so much agreement among the relevant speakers (in this case, professional philosophers) upon how to use a pair of terms with respect to an open class of sentences, then that pair of terms must mark *some* distinction; (2) Grice and Strawson argued (cf. Grice and Strawson, 1956) that the cases in which it appears that an analytic proposition was falsified can be explained away by contending that in each case the meaning of the words changed, and so the proposition that was at one time genuinely analytic was not the same proposition that was later falsified, although it was expressed by the very same sentence.

I agree with the first argument. There is an *obvious* difference (even if we have difficulty stating it) between, say, 'all bachelors are unmarried', as a representative analytic sentence, and 'my hat is on the table', as a representative synthetic sentence. It seems impossible to say that so obvious a distinction doesn't really have *any* basis. But Grice and Strawson's second argument seemed to me to be far less successful. Consider the statement that one cannot return to the place from which one started by travelling in a straight line in space in a constant direction. If this statement was once analytic or *a priori* (in 1951, few philosophers of an analytic persuasion would have troubled to distinguish the two notions), and was later falsified by the discovery (let us say) that our world is Riemannian in the large, then the Grice–Strawson rescue move would consist in saying that some term, say, 'straight line' has changed its meaning in the course of the change from Euclidean to Riemannian

cosmology. But even if 'straight line' has changed its 'connotations' – even if the theoretical aura surrounding the term is different – still this would not effect the *truth value* of the sentence unless the very reference of the term 'straight line' has changed, unless we are now referring to different paths in space as straight lines. But, having studied philosophy of physics and philosophy of geometry with Hans Reichenbach, I was not satisfied with this story at all. Whatever the nature of the conceptual revolution involved in the shift from Newtonian to relativistic cosmology may have been, it was not simply a matter of attaching the old labels, e.g. 'straight line', to new curves. What seemed *a priori* before the conceptual revolution was precisely that there *are* paths in space which behave in a Euclidean fashion; or, to drop reference to 'paths', what seemed *a priori* was precisely that there were infinitely many non-overlapping places (of, say, the size of an ordinary room) to get to. What turned out to be the case (or, rather, what will turn out to be the case if the universe in the large has compact spatial cross-sections), is precisely that there are only finitely many disjoint places (of the size of an ordinary room) in space to get to, travel as one will. Something literally *inconceivable* has turned out to be true; and it is not just a matter of attaching the old labels ('place', 'straight line') to different things.

To state the same point more abstractly: it often happens in a scientific revolution that something that was once taken to be an *a priori* truth is given up; and one cannot say that what has happened is simply that the words have been assigned to new referents, because, from the standpoint of the new theory, there are not and never were any objects which could plausibly have been the referents of the words in question. Nor can we say that the proposition in question used to mean that certain entities ('Euclidean straight lines', 'Euclidean places') *would* have certain properties if they existed, and that what has happened is that words ('straight line', 'place') which used to have no referents at all have now been assigned referents; for in the geometrical case there certainly were such entitites as *places the size of a room*, and what seemed necessary was that these *places* had the property of being infinite in number.

To put it another way, it seemed *a priori* that the terms 'path in space' and 'place the size of an ordinary room' had referents. To say that the existence propositions, 'There are places the size of an ordinary room' and 'There are paths in space', were *a posteriori* (in the old sense of the words), whereas the if-then proposition 'If anything is a place the size of a room, then there are infinitely many such places' is *a priori*, is utterly unmotivated, since these propositions did not differ in

epistemological or methodological status prior to the conceptual revolution under discussion.

I was driven to the conclusion that there was such a thing as the overthrow of a proposition that was once *a priori* (or that once had the status of what we *call* an '*a priori*' truth). If it could be rational to give up claims as self evident as the geometrical propositions just mentioned, then, it seemed to me that there was no basis for maintaining that there are any *absolutely a priori* truths, any truths that a rational man is *forbidden* to even doubt. Grice and Strawson were wrong; the overthrow of '*a priori*' propositions is not a mere illusion that can be explained away as change in the meaning of words. Quine's attack on the analytic–synthetic distinction, reconstrued as an attack on the *a priori–a posteriori* distinction, seemed to me to be correct. At the same time, if by an analytic truth one means a statement which is reducible to something like principles of elementary logic via meaning relations that are in some sense conventional, then it still seemed to me that there were analytic truths. Empiricist philosophers had bloated the analytic–synthetic distinction by making it coextensive with the *a priori–a posteriori* distinction; the question of the existence of analytic truths, in the sense just mentioned, had to be separated from the question whether any truths, even truths of elementary logic, were *a priori*.

In 'The analytic and the synthetic' I undertook the double task of defending Quine's insight with the aid of examples from the history of physics and geometry, and of clarifying the nature of the analytic–synthetic distinction itself. The conclusions I reached in the course of writing that paper had a far-reaching impact on my later views in the philosophy of mathematics, geometry, and quantum mechanics, as the reader can see by glancing at the papers in the first volume of these collected papers.

Conventionalism

An issue which is closely connected to the issues surrounding the analytic–synthetic distinction, and its misuse by philosophers, is the issue of conventionalism. Just as some philosophers try to clear up some philosophical puzzles by contending that certain statements which appear to be statements of fact are *really* 'analytic', so some philosophers contend that certain statements which appear to be statements of fact are really 'up for grabs', i.e. their truth-value is a matter of convention. Applications of this idea to the philosophy of language and to the philosophy of geometry are criticized in 'The refutation of conventionalism'. It is of interest that conventionalism in the philosophy of space and time was

originally motivated by a desire to give an account of the reference of scientific terms. Thus the critique of conventionalism naturally involves one in the very questions about reference that are taken up in the papers 'Explanation and reference' and 'The meaning of "meaning"'. (I also try to give an overview of relations between questions in the philosophy of language and questions in other parts of philosophy in the paper 'Language and reality'.)

I have not attempted in these papers to put forward any grand view of the nature of philosophy; nor do I have any such grand view to put forward if I would. It will be obvious that I do not agree with those who see philosophy as the history of 'howlers', and progress in philosophy as the debunking of howlers. It will also be obvious that I do not agree with those who see philosophy as the enterprise of putting forward *a priori* truths about the real world (since, for one thing, there are no *a priori* truths, in my view). I see philosophy as a field which has certain central questions, for example, the relation between thought and reality, and, to mention some questions about which I have *not* written, the relation between freedom and responsibility, and the nature of the good life. It seems obvious that in dealing with these questions philosophers have formulated rival research programs, that they have put forward general hypotheses, and that philosophers within each major research program have modified their hypotheses by trial and error, even if they sometimes refuse to admit that that is what they are doing. To that extent philosophy is a 'science'. To argue about whether philosophy is a science in any more serious sense seems to me to be hardly a useful occupation. The important thing is that in spite of the stereotypes of science and philosophy that have become blinkers inhibiting the view of laymen, scientists, and philosophers, science and philosophy are interdependent activities; philosophers have always found it essential to draw upon the scientific knowledge of the time, and scientists have always found it essential to do a certain amount of philosophy in their very scientific work, even if they denied that that was what they were doing. It does not seem to me important to decide whether science is philosophy or philosophy is science as long as one has a conception of both that makes both essential to a responsible view of the real world and of man's place in it.

Harvard University H.P.
September 1974