# Realism and Reason

# Philosophical Papers, Volume 3

### HILARY PUTNAM

Walter Beverly Pearson Professor of Mathematical Logic, Harvard University



Published by the Press Syndicate of the University of Cambridge The Pitt Building, Trumpington Street, Cambridge CB2 1RP 40 West 20th Street, New York, NY 10011-4211, USA 10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1983

First published 1983 First paperback edition 1985 Reprinted 1986, 1987, 1989, 1992, 1994, 1996

Library of Congress Catalog card number: 82-12903

A catalogue record for this book is available from the British Library

ISBN 0-521-24672-5 hardback ISBN 0-521-31394-5 paperback

Transferred to digital printing 2002

## Contents

Introduction		vii
1	Models and reality	I
2	Equivalence	26
3	Possibility and necessity	46
4	Reference and truth	69
5	'Two dogmas' revisited	87
6	There is at least one a priori truth	98
7	Analyticity and apriority: beyond Wittgenstein and	-
	Quine	115
8	Computational psychology and interpretation theory	139
9	Reflections on Goodman's Ways of Worldmaking	155
10	Convention: a theme in philosophy	170
ΙI	Philosophers and human understanding	184
12	Why there isn't a ready-made world	205
13	Why reason can't be naturalized	229
14	Quantum mechanics and the observer	248
15	Vagueness and alternative logic	271
16	Beyond historicism	287
Bib	liography	304
Acknowledgements		309
Index		310

## Introduction: An overview of the problem

The essays collected in this volume were written in a period of rethinking and reconsidering much of my philosophical position. A look at the introduction to the second volume of my *Philosophical Papers*, and at the essay titled 'Language and reality' in that volume, will reveal that in 1975 I thought that the errors and mistakes I detected in analytical philosophy were occasioned by 'naive verificationism' and 'sophisticated verificationism'. I described myself as a 'realist' (without any qualifying adjective), and I chiefly emphasized the importance of reference in determining meaning in opposition to the idea, traditional among both realists and idealists, that it is meaning that determines reference. Reference itself I described as a matter of causal connections. The following quotation (from 'Language and reality') illustrates these themes at work:

As language develops, the causal and noncausal links between bits of language and aspects of the world become more complex and more various. To look for any one uniform link between word or thought and object of word or thought is to look for the occult; but to see our evolving and expanding notion of reference as just a proliferating family is to miss the essence of the relation between language and reality. The essence of the relation is that language and thought do asymptotically correspond to reality, to some extent at least. A theory of reference is a theory of the correspondence in question. (p. 290)

Nothing in this quotation strikes me as exactly wrong even now; but what does strike me is that I was walking on a razor's edge without knowing it. In the intervening years I have come to see that one cannot come to grips with the real problems in philosophy without being more sensitive to the epistemological position of the philosopher than I was willing to be when I wrote those words. Becoming more sensitive to that position had consequences which I did not expect. It led me to think about questions which are thought to be more the province of 'continental philosophy' than of 'analytical philosophy', for instance, to think about the fact that our notions of rationality evolve in history (see, e.g., Putnam, 1981), and about the fact that one's own philosophical tradition has both a past and a

future. Some of the resulting reflections appear in this volume ('Convention: a theme in philosophy', and 'Beyond historicism', for example).

### The problems with the idea of truth as correspondence

The issue that first made me uncomfortable with my hard-line 'realist' position was one with which every philosopher is familiar: the notion that our words 'correspond' to determinate objects (where the notion of an 'object' is thought to have a determinate reference which is independent of conceptual scheme) had long seemed problematical, although I did not see any alternative to accepting it. By the time I had to give my Presidential Address (Putnam, 1976) to the Eastern Division of the American Philosophical Association in 1976 the problems with the notion had come to strike me as insuperable, and I finally began to search for an alternative view. What are those problems?

As far back as Berkeley and Kant it had been pointed out that the notion of a 'correspondence' is difficult once one becomes even a little bit psychologically sophisticated. If one is not psychologically sophisticated, then it appears easy to say how we 'put our words in correspondence with objects'. We teach a child a word, say 'table', by showing him the object and by using the word in various ways in the presence of that object (or, rather, kind of object) until the child comes to 'associate' the word with the object. In some sense, this is undeniably true. (As an ordinary language remark about, say, pedagogy, it is unproblematical; for, in such a context, the notions of an 'object' and of 'showing someone an object' and of 'associating a sound with a kind of object' are all taken for granted, as part of the linguistic background we assume.)

However, psychology is something that came on the scene at the same time as modern philosophy. Early philosophical psychologists – for example, Hume – pointed out that we do not literally have the object in our minds. The mind never compares an image or word with an object, but only with other images, words, beliefs, judgments, etc. The idea of a comparison of words or mental representations with objects is a senseless one. So how can a determinate correspondence between words or mental representations and external objects ever be singled out? How is the correspondence supposed to be fixed?

It is important to recognize that the problem does *not* depend upon the acceptance of any particular theory of mental phenomena such as Hume's theory of 'ideas and impressions', or the so-called 'sense

datum theory', or the 'Gestalt' theory, etc. On any theory, when the child learns the use of the word 'table', what happens is that the word is linked in certain complex ways ('associations') to certain mental phenomena (not all of them conscious, according to contemporary theory). Even if we replace mind talk with brain talk, and talk of 'mental phenomena' with talk of 'representations', 'information', computer programs, etc., as contemporary 'cognitive' psychologists do, the point remains: the functionally organized informationprocessing brain can manipulate 'perceptions' of tables, 'information' about tables, 'representations', etc., but not tables themselves. If we limit psychology, for the moment, to 'solipsistic' description, description of what happens in the individual considered in isolation from his environment, then no psychological facts in this narrow sense, no facts about introspectible mental phenomena (or even unconscious mental phenomena) and no facts about brain processing can fix any correspondence between a word or 'representation' and anything external to the mind or brain.

The only paper in this book that makes serious use of technical logic ('Models and reality') is not an attempt to solve this problem (how the 'correspondence' is fixed), but rather a verification that the problem really exists. What I show is that no matter what operational and theoretical constraints our practice may impose on our use of a language, there are always infinitely many different reference relations (different 'satisfaction relations', in the sense of formal semantics, or different correspondences) which satisfy all of the constraints.† A parable may explain the significance of this bit of logic.

The Parable: According to a famous passage in Either/Or, God is subject to recurrent boredom. (He created the world because He was bored; then He created Adam because He was bored with the world; then Adam was bored with Eve...) Not only is this correct, but, in fact, at the time of the Tower of Babel episode, God became bored again. Not only did He cause us to start speaking different languages, but He started to play around with the satisfaction relations, the 'correspondences', upon which the words—world connection depends.

To understand what He did, pretend that English was one of the languages in existence back then. Imagine that  $C_1$  and  $C_2$  are two

<sup>†</sup> The argument does not assume that we can know what operational and theoretical constraints our practice imposes on our use of the language, or know what constraints it would impose on that use if we knew all observational facts, or know which theory is the 'ideal theory': the theory, or set of theories that satisfy those constraints, as some of my critics have mistakenly taken it to do.

admissible 'correspondences' (satisfaction relations), i.e., that  $C_1$  (respectively,  $C_2$ ) is the satisfaction relation that one gets if  $M_1$  (respectively,  $M_2$ ) is the model that one uses to interpret English, where  $M_1$  and  $M_2$  are both models which satisfy all the operational and theoretical constraints that our practice imposes. Then what He did (to Hebrew, to Assyrian, to Coptic,...) was to specify that when a man used a word, the word would stand for its image or images under the correspondence  $C_1$ , and that when a woman used a word, the word would stand for its image or images under the correspondence  $C_2$ . †

This situation continues to the present day. Thus, there is one set of things – call it the set of cats – such that, when a man uses the word 'cat' it stands for that set (in a God's eye view), and a different set of things – call it the set of cats\* – such that when a woman uses the word 'cat' it stands for that set (in a God's eye view); there is a relation between events – call it the relation of causation – such that, when a man uses the word 'causes' it stands for that relation (in a God's eye view), and a different relation such that when a woman uses the word 'causes' it stands for that relation (in a God's eye view); and so on.

Notice that the same sentences are true under both of His reference-assignments, the sentences we accept generate the same experiential expectations under both schemes, the behavior that is associated with believing-true or desiring-true particular sentences is the same under both schemes, and if the expectations we have or the things we do are successful (respectively, unsuccessful) the sentences we are *then* required to accept by our operational and theoretical criteria are the same and their truth values are the same.

It amused God for a while to see men and women talking to each other, never noticing that they were almost never referring to the same objects, properties and relations, but then, once again, inevitably, He became bored, so He invented philosophers. (Here caution compels me to end the parable.)

There are a number of stock 'easy answers' to the problems of the determination of reference (most of them are discussed in 'Why there isn't a ready-made world'). Thus, a philosopher might say, 'When the child comes to "associate" the word *table* with certain perceptions, images, etc., he is "associating" it not in the semantical sense

<sup>†</sup> This was suggested by Bob Nozick. If you don't believe that it is operational and theoretical constraints that determine the class of 'true' sentences, then just let T be the class of true sentences, however that is determined, and let  $M_1$  and  $M_2$  be two extremely different models for T which agree on the interpretation of 'psychological' predicates (e.g., 'It looks to X exactly as if he is seeing a table') but not on the interpretation of 'external' predicates ('X is a table').

(table doesn't refer to the visual impressions which trigger the utterance there is a table in front of me, or whatever), but in a causal sense. He is caused to have certain beliefs, partly by the fact that certain visual impressions occur. But those visual impressions, "mental representations", or whatever, are in turn caused by certain external events. Normally, they are caused by the presence of a table, in fact. So, indirectly, the word table comes to be associated with external tables."

To see why this answer isn't a solution to the problem, imagine it being said first by a woman philosopher and then by a man philosopher. When the woman says this (now we are inside the parable again), she is pointing out that the child's belief there is a table in front of me is in a certain relation - the relation of effect\* - to certain visual impressions, and that these are in relation effect\* to certain external events. In fact they are caused\* by the presence\* of a table\*. So indirectly the word table has come to be associated with tables\*. When the man says this, he is pointing out that the same visual impressions are caused by the presence of a table. So the word table has come to be associated with tables. Of course they are both right. The word table is 'indirectly associated' with tables\* (in the way pointed out by the woman) and also 'indirectly associated' with tables (in the way pointed out by the man). It doesn't follow that there is such a thing as the correspondence (the One, metaphysically singled out correspondence) between words and things.

At this point in the dialogue, there is an argument that I invariably get from causal realists. This runs somewhat like this: 'You are caricaturing our position. A realist does not claim that reference is fixed by the conceptual connection (i.e., the connection in our theory) between the *terms* "reference", "causation", "sense impression", etc.; the realist claims that reference is fixed by causation itself.'

Here the philosopher is ignoring his own epistemological position. He is philosophizing as if naive realism were true of him (or, equivalently, as if he and he alone were in an absolute relation to the world). What he calls 'causation' really is causation, and of course there is a fixed, somehow singled-out, correspondence between the word and one definite relation in his case. Or so he assumes. But how this can be so was just the question at issue. (If this isn't clear, just imagine the words said first by a woman and then by a man, as before.)

A more sophisticated form of the same argument is this: 'Your argument only shows that reference is not fixed by anything psychological, anything "inside the head". But that is no problem: why can't reference be fixed by something non-psychological?'

The answer, quite simply, is that the idea that the 'nonpsychological' fixes reference - i.e., that nature itself determines what our words stand for - is totally unintelligible. At bottom, to think that a sign-relation is built into nature is to revert to medieval essentialism, to the idea that there are 'self-identifying objects' and 'species' out there. (This is discussed at length in 'Why there isn't a ready-made world'.) Such an idea made sense in the context of a medieval world view, which had not only an elaborate ontology (essence and existence, substantial form, etc.) to back it up, but also an elaborate psychology (e.g., Aquinas' distinction between the 'passive imagination' and the 'productive imagination' as well as his 'phantasms', 'intellectual species', etc.) and an elaborate correspondence between the two (God had arranged it so that the 'intellectual species' produced by the productive imagination acting on the phantasms would have a pre-established correspondence to the substantial forms). In the context of a twentieth-century world view, by contrast, to say in one's most intimidating tone of voice 'I believe that causal connections determine what our words correspond to' is only to say that one believes in a one-knows-not-what which solves our problem one-knows-not-how.

### Ontological relativity?

A solution that has occurred to some philosophers is to keep the idea of 'correspondence', but simply to abandon the idea that there is one correspondence that is 'fixed', one intended reference-relation. If we take this line, then the word 'table' refers to one definite set of things in an ordinary 'empirical' sense, but not in a 'transcendental' sense (not from the standpoint of the metalanguage). When we say that there is one set which is the set of tables and a different set which is the set of chairs, that statement is true (construed as an ordinary first-order statement). There is just one set which is the set of chairs and just one set which is the set of tables in each model, and the set of tables in a model is a different set from the set of chairs in that same model. But that does not mean that there is one set which is the set of chairs in every admissible model. The set of chairs in one model may be a subset or a superset of the set of tables in a different model. How we imagine tables and chairs, what experiences we have when we see and touch tables and chairs, what we do in their presence, etc., are all unaffected by the lack of a unique assignment of objects and sets of objects to our words; the words 'see', 'touch', 'sit down', etc., simply change their reference from model to model in such a way that nothing we can notice is ever affected.

The doctrine just described has been called 'ontological relativity'. It was suggested by Quine (generally he suggests this is the stance to take to all languages other than one's own, although there are places where he speaks of the reference of even his own terms as 'free floating'.) The doctrine that Davidson calls 'the empty reference theory' is, perhaps, the same doctrine.

This doctrine cannot, however, be accepted. I cannot accept it for my own language, because to do so would turn the notion of an object into a totally metaphysical notion. I know what tables are and what cats are and what black holes are. But what am I to make of the notion of an X which is a table or a cat or a black hole (or the number three or...)? An object which has no properties at all in itself and any property you like 'in a model' is an inconceivable Ding an sich. The doctrine of ontological relativity avoids the problems of medieval philosophy (the problems of classical realism) but it takes on the problems of Kantian metaphysics in their place. Nor can the doctrine be accepted for languages other than my own; the human situation is symmetrical. If other people's words do not determinately refer, then neither do my own. (Quine's view is discussed in 'Why reason can't be naturalized'.)

### Disquotational theories of truth and reference

If the picture of the language user that we have thus far discussed - the picture of (there being) one particular correspondence between what is 'inside' the mind or brain (include language) and what is 'outside' - leads to the metaphysical fantasy of a 'ready-made world', with self-identifying objects, 'built-in' structure, essences, or whatever, and the modified picture of the mind or brain simply accepting a whole lot of different correspondences, without trying to 'fix' any particular one as the intended correspondence between word and object, leads to the metaphysical fantasy of a 'noumenal' world, with no determinate relation to our experiential world, then the trouble with this entire discussion must lie at a deeper level. It must lie, in fact, in the common assumption of both pictures: that we understand such notions as 'refers to' and 'corresponds to' by associating these notions with Platonic objects ('correspondences') either unique objects or else whole batches of objects. Once this assumption has been made (usually uncritically and inexplicitly) then the entire system of competing philosophical theories and arguments unwinds itself with a sort of inevitability. But can we avoid this common assumption?

One attempt to do this is the 'disquotational' theory of truth. On this theory, we understand the word 'true' not by associating that word with a property, or a correspondence, but by learning such facts as the obvious fact that

(1) 'Snow is white' is true if and only if snow is white.

Learning such 'T-sentences' (sentences of the form "P" is true if and only if P') is (in a rational reconstruction, at least) the process of coming to understand the word 'true'. Since associating the word 'true' with an 'object' (a property or correspondence) need not be involved at any stage in acquiring these facts, the whole idea of 'correspondence' is misleading. (Alternatively: (1) does say, in a way, that 'Snow is white' is true if and only if it corresponds to what is the case; so the disquotation theory is the correspondence theory, 'properly understood'.)

An objection to the disquotation theory which is still occasionally heard is that the understanding of the truth-functional connective 'if and only if' presupposes the very notions of truth and falsity which are being explained. This objection misses the thrust of the theory, however: the theory is not that we understand 'true' by learning that "Snow is white" is true' is true if and only if it is true that snow is white, but that we understand 'true' by being trained to assert "Snow is white" is true' when (and only when) we are prepared to assert 'Snow is white', and similarly in similar cases. The disquotational view is at home in a larger view on which our understanding of our first language comes about through the internalization of assertibility conditions and not through the learning of truth conditions in the realist sense.

If such a theory is to have any explanatory power, then something must be said about the notion of assertion. It seems to me that the theory does not work at all if assertion is taken in a 'thin' (or merely behavioral) sense, as it is, for example, by Quine. If 'asserting' p is merely uttering p (or subvocalizing 'yes' when another utters p), then all the theory tells us is that we are disposed to utter the noise '... is true' when we are prepared to utter certain other noises. If assertion is to be taken in a suitably 'thick' sense, however, then we have to recognize that asserting is guided by notions of correctness and incorrectness. But then the problem of truth reappears when we ask for an account of what it is for an assertion to be correct and what it is for it to be incorrect.

But why shouldn't the disquotationalist reply that 'correct' is just a synonym for 'true', and that he has already explained how 'true'

works? A comparison with the case of an ordinary first-order scientific sentence, say, 'there is electric current flowing through this wire' may clarify the issue. An assertibility condition (a probabilistic one, within a scientific theory) for the sentence 'there is electric current flowing through this wire' is that the needle of a suitably attached voltmeter be deflected. But describing, as well as one can, what the assertibility conditions for this sentence are does not preempt the question 'What is electric current?'. Answering that question calls for a theory of electricity, not for 'transcendental' remarks about assertibility conditions for 'electricity talk'. Similarly, I suggest, describing assertibility conditions for 'This sentence is true' (or 'That is right', or 'That is correct', etc.) does not preempt the question 'What is the nature of truth?' (or rightness, or correctness etc.). If a philosopher says that truth is different from electricity in precisely this way: that there is room for a theory of electricity but no room for a theory of truth, that knowing the assertibility conditions is knowing all there is to know about truth, then, in so far as I understand him at all, he is denying that there is a property of truth (or a property of rightness, or correctness), not just in the realist sense, but in any sense. But this is to deny that our thoughts and assertions are thoughts and assertions. (This is discussed in 'Vagueness and alternative logic' and also figures in my criticism of naturalized epistemology in 'Why reason can't be naturalized'.)

A disquotational theory of reference holds that we understand 'refers to' not by associating the phrase 'refers to' with a 'correspondence', but by learning such assertibility conditions as the following:

(2) 'Cat' refers to an object X if and only if X is a cat.

Interpreted as an assertibility condition, what (2) tells us is to assert 'That sentence refers to (contains a word which refers to) cats' when and only when a sentence has been used which contains the word 'cat' or some word W such that one is prepared to assert

(3) Something is a W if and only if it is a cat.

Once again, the view finds its home in a larger view according to which the understanding of our language is through the internalization of assertibility conditions, and not through the learning of truth conditions in the realist sense. And once again, the tenability of the view will depend upon the availability of a sufficiently substantial notion of assertibility.

### The views of Michael Dummett

If neither a correspondence theory nor a pure disquotational theory of truth is of much help, the situation is reminiscent of a common predicament in philosophy. As Strawson remarked many years ago, we are constantly being asked to choose between *metaphysical* positions on the one hand and *reductionist* positions on the other, and what is terribly difficult (but what makes the game of philosophy worth the candle) is to show that the metaphysical mystery is not the only alternative to the simplistic position of the reductionist (and, of course, vice versa).

During the years that I was wrestling with the problems I have just described, the first clear indication that a coherent alternative to both the correspondence theory and the pure disquotational theory might be available came from the writings of Michael Dummett. Dummett considers the learning of a language to be the learning of a practice and not of a set of correspondences; he considers the speaker's knowledge of his native language to consist in the implicit knowledge of the conditions under which the sentences of that language are assertible (a sort of recognition ability); but he rejects the physicalist identification of asserting with uttering, or with uttering plus a particular schedule of conditioning or a particular causal history. Rather, he identifies knowing when a sentence is assertible with knowing when it would be justified.

The use of the word 'true' is not, on this theory, a mere sign that a sentence is being 'reaffirmed', as it is on the disquotation theory. To be true is to be justified. Reference, however, is not something prior to truth; rather, knowing the conditions under which sentences about, say, tables, are true is knowing what 'table' refers to (as on a disquotational theory of reference). Indeed this idea – that objects and reference arise out of discourse rather than being prior to discourse – is rather widespread in twentieth-century philosophy, in both its analytical and 'continental' varieties. (Gadamer, for example, speaks of objects as 'emerging' from discourse.)

### Truth as justification

The formula 'truth is justification' is misleading in a number of ways, however, which is why I have avoided it in my own writings, in spite of the inspiration I received from Dummett's work. For one thing, it suggests something which Dummett indeed believes and I do not (see 'Reference and truth'): that one can specify in an effective way

what the justification conditions for all the sentences of a natural language are. Secondly, it suggests something on which Dummett's writing is rather ambiguous: that there is such a thing as *conclusive* justification, even in the case of empirical sentences.

My own view (for which I have suggested the name 'internal realism') is that truth is to be identified with justification in the sense of *idealized* justification, as opposed to justification-on-present-evidence.† Sometimes this seems to be Dummett's view too (at such times he speaks of a 'gap between justification and truth'); at other times he writes as if ordinary-language-sentences about material objects outside of theoretical science could be conclusively verified.

Consider the sentence 'There is a chair in my office right now.' Under sufficiently good epistemic conditions any normal person could verify this, where sufficiently good epistemic conditions might, for example, consist in one's having good vision, being in my office now with the light on, not having taken a hallucinogenic agent, etc. How do I know these are better conditions for this sort of judgment than conditions under which one does not have very good vision, or in which one is looking into the room through a telescope from a great distance, or conditions in which one has taken LSD? Partly by knowing how talk of this sort operates (what the 'language game' is, in Wittgenstein's sense), and partly by having a lot of empirical information. There is no single general rule or universal method for knowing what conditions are better or worse for justifying an arbitrary empirical judgment.

On this view (mine), then, 'truth' (idealized justification) is as vague, interest relative, and context sensitive as we are. The 'truth conditions' for an arbitrary sentence are not surveyable in Dummett's sense. I reject 'meaning theories'.

If truth conditions and assertibility conditions are not surveyable,

<sup>†</sup> If one accepts this account of truth, then (as Michael Dummett has pointed out in connection with his, related, account) the question of the status of some of the traditional laws of logic must be reopened, notably that of the status of the principle of bivalence (every statement is determinately true or false). This connection between the philosophy of language and the choice of a logic is a fascinating one; it is commented on in two of the papers in this volume, 'Vagueness and alternative logic', and 'Quantum mechanics and the observer'. The wider question of the 'apriority' of the laws of logic is also discussed in a series of papers in this volume.

<sup>†</sup> Of course it is possible to survey the surface truth conditions for the sentences of a language (if the language has been canonically formalized), e.g., 'snow is white' is true if and only if snow is white, and even to write down a theory which has them all as theorems. This is what Donald Davidson calls a 'meaning theory' for the language. Where I differ from Davidson is in not attaching any philosophical significance to the surface truth conditions in the absence of an account of what it is for a native speaker to know them. (Cf. 'Reference and truth'.)

how do we learn them? We learn them just the way Dummett thinks, at least in the case of the less theoretical parts of the language, by acquiring a practice. What Dummett misses, in my view, is that what we acquire is not a knowledge that can be applied as if it were an algorithm. We do learn that in certain circumstances we are supposed to accept 'There is a chair in front of me' (normally). But we are expected to use our heads. We can refuse to accept 'There is a chair in front of me' even when it looks to us exactly as if there is a chair in front of us, if our general intelligence screams 'override'. The impossibility (in practice at least) of formalizing the assertibility conditions for arbitrary sentences is just the impossibility of formalizing general intelligence itself.

If assertibility (in the sense of warranted assertibility) is not formalizable, idealized warranted assertibility (truth) is even less so, for the notion of better and worse epistemic conditions (for a particular judgment) upon which it depends is revisable as our empirical knowledge increases. That it is, nevertheless, a meaningful notion; that there are better and worse epistemic conditions for most judgments, and a fact of the matter as to what the verdict would be if the conditions were sufficiently good, a verdict to which opinion would 'converge' if we were reasonable, is the heart of my own 'realism'. It is a kind of realism, and I mean it to be a human kind of realism, a belief that there is a fact of the matter as to what is rightly assertible for us, as opposed to what is rightly assertible from the God's eye view so dear to the classical metaphysical realist. (On this, see 'Reflections on Goodman's Ways of Worldmaking'.)