RENEWING PHILOSOPHY

HILARY PUTNAM

RENEWING PHILOSOPHY

Hilary Putnam

HARVARD UNIVERSITY PRESS

Cambridge, Massachusetts

London, England

Printed in the United States of America
FOURTH PRINTING, 1995

Library of Congress Cataloging-in-Publication Data
Putnam, Hilary.
Renewing philosophy / Hilary Putnam.
p. cm.
Includes bibliographical references and index.
ISBN 0-674-76093-X (alk. paper) (cloth)
ISBN 0-674-76094-8 (pbk.)
1. Philosophy. 2. Philosophy and science. I. Title.
B29.P88 1992
100—dc20 92-10854
CIP

Preface

The present book grew out of the Gifford Lectures, which I delivered at the University of St. Andrews in the Fall of 1990, and, with one exception, its chapters are quite close to the lectures as given. (Chapter 5 has been very substantially re-written. In addition, there was an opening lecture in which, perhaps perversely, I chose to deal with the present situation in quantum mechanics and its philosophical significance, which I decided did not really belong with the others.)

At first blush, the topics with which the lectures dealt may seem to have little relation to one another: I spoke of reference and realism and religion and even of the foundations of democratic politics. Yet my choice of these topics was not an arbitrary one. I was guided, of course, by my own past areas of concern, since it would have been foolish to lecture on topics on which I had not done serious thinking and writing in the past, but beyond that I was guided by a conviction that the present situation in philosophy is one that calls for a revitalization, a renewal, of the subject. Thus this book, in addition to addressing several topics individually, offers a diagnosis of the present situation in philosophy as a whole and suggests the directions in which we might look for such a renewal. That suggestion does not take the form of a manifesto, however, but rather takes the form of a series of reflections on various philosophical ideas.

Analytic philosophy has become increasingly dominated by

the idea that science, and only science, describes the world as it is in itself, independent of perspective. To be sure, there are within analytic philosophy important figures who combat this scientism: one has only to mention Peter Strawson, or Saul Kripke, or John McDowell, or Michael Dummett. Nevertheless, the idea that science leaves no room for an independent philosophical enterprise has reached the point at which leading practitioners sometimes suggest that all that is left for philosophy is to try to anticipate what the presumed scientific solutions to all metaphysical problems will eventually look like. (This is accompanied by the weird belief that one *can* anticipate that on the basis of *present-day* science!) The first three chapters in this volume are concerned to show that there is extremely little to this idea. I begin with a look at some of the ways in which philosophers have suggested that modern science explains the link between language and the world. The first chapter discusses the decidedly premature enthusiasm that some philosophers feel for "Artificial Intelligence". The second chapter takes on the idea that evolutionary theory is the key to the phenomenon of representation, while the third chapter subjects to close scrutiny a contemporary philosopher's claim that one can define reference in terms of causality. I try to show that these ideas lack scientific and philosophical substance, while gaining prestige from the general philosophical climate of deference to the supposed metaphysical significance of science.

Perhaps the most impressive case for the view that one *should* look to present-day science, and especially to physics, for at least a very good sketch of an adequate metaphysics has been made by the British philosopher Bernard Williams, and after a chapter which deals with some of the problems faced by both relativistic and materialistic metaphysicians, I devote a chapter to a close examination of his views.

Not all present-day philosophers are overawed by science,

however, and some of the philosophers who are not—philosophers like Derrida, or, in the English-speaking world, Nelson Goodman or Richard Rorty—have reacted to the difficulty of making sense of our cognitive relation to the world by denying that we do have a cognitive relation to extralinguistic reality. In my sixth chapter, I criticize these thinkers for throwing away the baby with the bathwater. In the seventh and eighth chapters, I examine Wittgenstein's "Lectures on Religious Belief", arguing that those lectures demonstrate how a philosopher can lead us to see our various forms of life differently without being either scientistic or irresponsibly metaphysical, while in the concluding chapter I try to show how John Dewey's political philosophy exhibits the same possibility in a very different way.

The two months that I spent at St. Andrews giving these lectures were a sheer delight, and I profited more than I can say from the companionship and the philosophical conversation of the remarkable group of brilliant and dedicated philosophers there, particularly Peter Clark, Bob Hale, John Haldane, Stephen Read, Leslie Stevenson, John Skorupski, and Crispin Wright. As always in recent years, many of the ideas in these chapters were first tried out in conversation with Jim Conant, and Chapter 5, in particular, owes a great deal to those conversations. Chapter 9 first appeared, in a slightly different form, in *Southern California Law Review* 63 (1990): 1671–97, and is reprinted here with that journal's permission. I am also grateful to Bengt Molander of the University of Uppsala and to Ben-Ami Sharfstein of the University of Tel Aviv, both of whom read earlier versions and made valuable suggestions. At a very late stage, excellent suggestions were also made by the referees for the Harvard University Press, not all of which I could take up without changing the character of the work, but some of which I have responded to, and some of which will show their effect in my future writing. The most valuable suggestions of

all were made by Ruth Anna Putnam, who provided not only the affection and support which mean so much, but whose close reading and fine criticism certainly made this a much better book.

RENEWING PHILOSOPHY

Contents

	Preface	ix
1.	The Project of Artificial Intelligence	1
2.	Does Evolution Explain Representation?	19
3.	A Theory of Reference	35
4.	Materialism and Relativism	60
5.	Bernard Williams and the Absolute Conception of the World	80
6.	Irrealism and Deconstruction	108
7.	Wittgenstein on Religious Belief	134
8.	Wittgenstein on Reference and Relativism	158
9.	A Reconsideration of Deweyan Democracy	180
	Notes	203
	Index	227

1

The Project of Artificial Intelligence

Traditionally Gifford Lectures have dealt with questions connected with religion. In recent years, although reference to religion has never been wholly absent, they have sometimes been given by scientists and philosophers of science, and have dealt with the latest knowledge in cosmology, elementary particle physics, and so on. No doubt the change reflects a change in the culture, and particularly in the philosophical culture. But these facts about the Gifford Lectures—their historical concern with religion and their more recent concern with science—both speak to me. As a practicing Jew, I am someone for whom the religious dimension of life has become increasingly important, although it is not a dimension that I know how to philosophize about except by indirection; and the study of science has loomed large in my life. In fact, when I first began to teach philosophy, back in the early 1950s, I thought of myself as a philosopher of science (although I included philosophy of language and philosophy of mind in my generous interpretation of the phrase "philosophy of science"). Those who know my writings from that period may wonder how I reconciled my religious streak, which existed to some extent even back then, and my general scientific materialist worldview at that time. The answer is that I didn't reconcile them. I was a thoroughgoing atheist, and I was a believer. I simply kept these two parts of myself separate.

2

In the main, however, it was the scientific materialist that was dominant in me in the fifties and sixties. I believed that everything there is can be explained and described by a single theory. Of course we shall never know that theory in detail, and even about the general principles we shall always be somewhat in error. But I believed that we can see in present-day science what the general outlines of such a theory must look like. In particular, I believed that the best metaphysics is physics, or, more precisely, that the best metaphysics is what the positivists called "unified science", science pictured as based on and unified by the application of the laws of fundamental physics. In our time, Bernard Williams has claimed that we have at least a sketch of an "absolute conception of the world" in present-day physics.[1] Many analytic philosophers today subscribe to such a view, and for a philosopher who subscribes to it the task of philosophy becomes largely one of commenting on and speculating about the progress of science, especially as it bears or seems to bear on the various traditional problems of philosophy.

When I was young, a very different conception of philosophy was represented by the work of John Dewey. Dewey held that the idea of a single theory that explains everything has been a disaster in the history of philosophy. Science itself, Dewey once pointed out, has never consisted of a single unified theory, nor have the various theories which existed at any one time ever been wholly consistent. While we should not stop trying to make our theories consistent—Dewey did not regard inconsistency as a *virtue*—in philosophy we should abandon the dream of a single absolute conception of the world, he thought. Instead of seeking a final theory—whether it calls itself an "absolute conception of the world" or not—that would explain everything, we should see philosophy as a reflection on how human beings can resolve the various sorts of "problematical situations" that

they encounter, whether in science, in ethics, in politics, in education, or wherever. My own philosophical evolution has been from a view like Bernard Williams' to a view much more like John Dewey's. In this book I want to explain and, to the extent possible in the space available, to justify this change in my philosophical attitude.

In the first three chapters, I begin with a look at some of the ways in which philosophers have suggested that modern cognitive science explains the the link between language and the world. This chapter deals with Artificial Intelligence. Chapter 2 will discuss the idea that evolutionary theory is the key to the mysteries of intentionality (i.e., of truth and reference), while Chapter 3 will discuss the claim made by the philosopher Jerry Fodor that one can define reference in terms of causal/counterfactual notions. In particular, I want to suggest that we can and should accept the idea that cognitive psychology does not simply reduce to brain science *cum* computer science, in the way that so many people (including most practitioners of "cognitive science") expect it to.

I just spoke of a particular picture of what the scientific worldview is, the view that science ultimately reduces to physics, or at least is unified by the world picture of physics. The idea of the mind as a sort of "reckoning machine" goes back to the birth of that "scientific worldview" in the seventeenth and eighteenth centuries. For example, Hobbes suggested that thinking is appropriately called "reckoning", because it really is a manipulation of signs according to rules (analogous to calculating rules), and La Mettrie scandalized his time with the claim that man is just a machine *(L'Homme Machine)*.[2] These ideas were, not surprisingly, associated with materialism. And the question which anyone who touches on the topic of Artificial Intelligence is asked again and again is "Do you think that a computing machine could have intelligence, conscious-

ness, and so on, in the way that human beings do?" Sometimes the question is meant as "could it in principle" and sometimes as "could it really, in practice" (to my mind, the far more interesting question).

The story of the computer, and of Alan Turing's role in the conception of the modern computer, has often been told. In the thirties, Turing formulated the notion of computability[3] in terms which connect directly with computers (which had not yet been invented). In fact, the modern digital computer is a realization of the idea of a "universal Turing machine". A couple of decades later materialists like my former self came to claim that "the mind is a Turing machine". It is interesting to ask why this seemed so evident to me (and still seems evident to many philosophers of mind).

If the whole human body is a physical system obeying the laws of Newtonian physics, and if any such system, up to and including the whole physical universe, is at least metaphorically a machine, then the whole human body is at least metaphorically a machine. And materialists believe that a human being *is* just a living human body. So, as long as they assume that quantum mechanics cannot be relevant to the philosophy of mind (as I did when I made this suggestion),[4] materialists are committed to the view that a human being is—at least metaphorically—a machine. It is understandable that the notion of a Turing machine might be seen as just a way of making this materialist idea precise. Understandable, but hardly well thought out.

The problem is the following: a "machine" in the sense of a physical system obeying the laws of Newtonian physics need not be a Turing machine. (In defense of my former views, I should say that this was not known in the early 1960s when I proposed my so-called functionalist account of mind.) For a Turing machine can compute a function only if that function

belongs to a certain class of functions, the so-called general recursive functions. But it has been proved that there exist possible physical systems whose time evolution is not describable by a recursive function, even when the initial condition of the system is so describable. (The wave equation of classical physics has been shown to give rise to examples.) In less technical language, what this means is that there exist physically possible analogue devices which can "compute" non-recursive functions.[5] Even if such devices cannot actually be prepared by a physicist (and Georg Kreisel has pointed out that no theorem has been proved *excluding* the preparation of such a device),[6] it does not follow that they do not occur in nature. Moreover, there is no reason at all why the real numbers describing the condition at a specified time of a naturally occurring physical system should be "recursive". So, for more than one reason, a naturally occurring physical system might well have a trajectory which "computed" a non-recursive function.

You may wonder, then, why I assumed that a human being could be, at least as a reasonable idealization, regarded as a Turing machine. One reason was that the following bit of reasoning occurred to me. A human being cannot live forever. A human being is finite in space and time. And the words and actions—the "outputs", in computer jargon—of a human being, insofar as they are perceivable by the unaided senses of other human beings (and we might plausibly assume that this is the level of accuracy aimed at in cognitive psychology) can be described by physical parameters which are specified to only a certain macroscopic level of accuracy. But this means that the "outputs" can be predicted during the finite time the human lives by a sufficiently good approximation to the actual continuous trajectory, and such a "sufficiently good approximation" can be a recursive function. (Any function can be approximated to any fixed level of accuracy by a recursive function over any

finite time interval.) Since we may assume that the possible values of the boundary parameters are also restricted to a finite range, a finite set of such recursive functions will give the behavior of the human being under all possible conditions in the specified range to the desired accuracy. (Since the laws of motion are continuous, the boundary conditions need only to be known to within some appropriate Δ in order to predict the trajectory of the system to within the specified accuracy.) But if that is the case, the "outputs"—what the human says and does—can be predicted by a Turing machine. (In fact, the Turing machine only has to compute the values of whichever recursive function in the finite set corresponds to the values that the boundary conditions have taken on), and such a Turing machine could, in principle, simulate the behavior in question as well as predict it.

This argument proves too much and too little, however. On the one hand, it proves that *every* physical system whose behavior we want to know only up to some specified level of accuracy and whose "lifetime" is finite can be simulated by an automaton! But it does not prove that such a simulation is in any sense a *perspicuous representation* of the behavior of the system. When an airplane is flying through the air at less than supersonic speeds, it is perspicuous to represent the air as a continuous liquid, and *not* as an automaton. On the other hand it proves too little from the point of view of those who want to say that the real value of computational models is that they show what our "competence" is in idealization from such limitations as the finiteness of our memory or our lifetimes. According to such thinkers,[7] *if we were able to live forever, and were allowed access to a potentially infinite memory storage, still all our linguistic behavior could be simulated by an automaton*. We are best "idealized" as Turing machines, such thinkers

say, when what is at stake is not our actual "performance" but our "competence". Since the proof of the little theorem I just demonstrated depended *essentially* on assuming that we do not live forever and on assuming that the boundary conditions have a finite range (which excludes a potentially infinite external memory), it offers no comfort to such a point of view.

Again, it might be said that any non-recursivities either in our initial conditions or in our space-time trajectories could not be reliably detected and hence would have no "cognitive" significance. But it is one thing to claim that the *particular* non-recursive function a human might compute if the human (under a certain idealization) were allowed to live forever has no cognitive significance, and another to say that the whole infinite trajectory can *therefore* be approximated by a Turing machine. Needless to say, what follows the "therefore" in this last sentence does not follow logically from the antecedent! (Recall how in the "chaos" phenomena small perturbations become magnified in the course of time.)

In sum, it does not seem that there is any principled reason why we must be perspicuously representable as Turing machines, *even assuming the truth of materialism*. (Or any reason why we must be representable in this way at all—even non-perspicuously—under the idealization that we live forever and have potentially infinite external memories). That is all I shall say about the question whether we are (or can be represented as) Turing machines "in principle".

On the other hand, the interesting question *is* precisely whether we are perspicuously representable as Turing machines, even if there are no a priori answers to be had to this question. And this is something that can be found out only by seeing if we can "simulate" human intelligence *in practice*. Accordingly, it is to this question that I now turn.

8

Induction and Artificial Intelligence

A central part of human intelligence is the ability to make inductive inferences, that is, to learn from experience. In the case of deductive logic, we have discovered a set of rules which satisfactorily formalize valid inference. In the case of inductive logic this has not so far proved possible, and it is worthwhile pausing to ask why.

In the first place, it is not clear just how large the scope of inductive logic is supposed to be. Some writers consider the "hypothetico-deductive method"—that is, the inference from the success of a theory's predictions to the acceptability of the theory—the most important part of inductive logic, while others regard it as already belonging to a different subject. Of course, if by induction we mean "any method of valid inference which is not deductive", then the scope of the topic of inductive logic will be simply enormous.

If the success of a large number of predictions—say, a thousand, or ten thousand—which are not themselves consequences of the auxiliary hypotheses alone *always* confirmed a theory, then the hypothetico-deductive inference, at least, would be easy to formalize. But problems arise at once. Some theories are accepted when the number of confirmed predictions is still very small—this was the case with the general theory of relativity, for example. To take care of such cases, we postulate that it is not only the number of confirmed predictions that matters, but also the elegance or simplicity of the theory: but can such quasi-aesthetic notions as "elegance" and "simplicity" really be formalized? Formal measures have indeed been proposed, but it cannot be said that they shed any light on real-life scientific inference. Moreover, a confirmed theory sometimes fits badly with background knowledge; in some cases, we conclude the theory cannot be true, while in others we conclude that the