MACHINES AND INTELLIGENCE

A Critique of Arguments Against the Possibility of Artificial Intelligence

STUART GOLDKIND

MACHINES AND INTELLIGENCE

A Critique of Arguments Against the Possibility of Artificial Intelligence

STUART GOLDKIND

Contributions to the Study of Computer Science, Number 2

GREENWOOD PRESS
NEW YORK • WESTPORT, CONNECTICUT • LONDON

Library of Congress Cataloging-in-Publication Data

Goldkind, Stuart, 1950-
Machines and intelligence.

(Contributions to the study of computer science,
ISSN 0734-757X ; no. 2)
Bibliography: p.
Includes index.
1. Artificial intelligence. I. Title. II. Series.
Q335.G65 1987 006.3 86-25712
ISBN 0-313-25450-8 (lib. bdg. : alk. paper)

Library of Congress Catalog Card Number: 86-25712
ISBN: 0-313-25450-8
ISSN: 0734-757X

First published in 1987

Greenwood Press, Inc.
88 Post Road West, Westport, Connecticut 06881

Printed in the United States of America

The paper used in this book complies with the Permanent Paper Standard issued by the National Information Standards Organization (Z39.48-1984).

10 9 8 7 6 5 4 3 2 1

Copyright Acknowledgments

The author gratefully acknowledges permission to use portions of the following copyrighted material.

Alan Turing, "Computing Machinery and Intelligence," *Mind*, Vol. LIX, no. 236 (1950). Reprinted with permission.

Richard Taylor, *Action and Purpose* (Englewood Cliffs, N.J.: Prentice-Hall, 1966). Reprinted with permission.

Hubert L. Dreyfus, *What Computers Can't Do: The Limits of Artificial Intelligence* (New York: Harper & Row, 1979). Copyright © 1972, 1979 by Hubert L. Dreyfus.

MACHINES AND INTELLIGENCE

Recent Titles in
Contributions to the Study of Computer Science

Reckoners: The Prehistory of the Digital Computer, From Relays to the Stored Program Concept, 1935-1945
Paul E. Ceruzzi

PREFACE

One of the central claims of mechanism has been that human beings are nothing more than (very complicated) machines. This claim has been found quite disturbing by some people, and over the years there have been numerous attempts to remove the annoyance by providing a convincing refutation. The most familiar form that such attempts at refutation take is that of an argument purporting to show that humans are possessed of some quality, feature, characteristic, or ability, X, which machines cannot possibly have. (For example, in arguments based on some famous results of Kurt Godel, the claim is made that machines by their very nature cannot produce certain theorems.) The evidence for humans possessing X, on the other hand, is usually empirical: we have merely to look in order to see that humans have X. The possession of X, then, is supposed to distinguish men from machines.

My own feeling is that none of these attempts will ever succeed; I will not, however, argue for this conclusion anywhere in the present work. This is because I see no way of proving the quite general statement that there is no such trait X. Instead, I choose to examine and criticise some ten or twelve specific arguments, each of which advances

some particular candidate for X. In each case we will find that the argument in question fails to establish any essential difference between men and machines.

Chapter 1 is the exception to the general strategy of the rest of the work in that it contains no candidate for a trait, X, possessed by humans but not by machines. However, the question treated there (the nature and validity of the Turing Test) is too fundamental to the whole area of the man/machine controversy to be neglected.

Chapter 2 deals with a well known book, *What Computers Can't Do*, by Hubert Dreyfus (Dreyfus, 1979). Dreyfus applies his background in phenomenology to the question of whether work in artificial intelligence can ever succeed. We will examine seven arguments which Dreyfus presents against machines being capable of such theings, for example, as chess playing, natural language understanding, and the ability to deal with contexts.

Chapter 3 deals with an argument devised by the present author concerning a problem raised by Alan Turing. This argument seeks to show that machines, because of their "rigid," "programmed" nature, cannot make the sorts of mistakes of which humans are capable. In this chapter we find that machines can indeed to something very much like what humans do when they make mistakes.

Chapter 4 is concerned with two arguments from Richard Taylor's book *Action and Purpose* (Taylor, 1967). There Taylor seeks to show that machines are incapable of purposive behavior. We will find that his arguments, while true of some machines, are not sufficiently general to sustain a conclusion about all machines.

Taking off from a discussion of Norman Malcolm's article, "The Conceivability of Mechanism" (Malcolm, 1968), Chapter 5 explores the relationship between causal and purposive explanations of behavior. In that chapter it is seen that a plausible argument against machines being able to behave purposefully cannot be supported by a supposed incompatibility between the two types of explanation.

Of course it is not possible to discuss every argument that has ever been advanced against the possibility of artificial intelligence; in a work of this size, in fact, it is not even possible to cover all of the most prominent arguments. What I have tried to do is to deal with arguments

which, for one reason or another, have not received adequate treatment in the past. For a definitive treatment of the much discussed arguments based on Godel's results (and an index into the all too extensive literature surrounding those arguments), see "The Abilities of Men and Machines" in (Dennett, 1978). And for a reply to Searle's "Chinese Room" arguments, along with a similar index into the literature, see (Hofstadter and Dennett, 1981) where Searle's original article is also reproduced. As for the rest, the usual apologies are extended to any who might feel that something of importance has been left out. I would welcome any communications along these lines (because of the possibility of revising the present text for future editions), or on other matters pertaining to this work; such communications may be sent to me in care of the publishers.

Finally, a note on the purpose of this enterprise: Why concern ourselves with these sorts of arguments? It is important to realize that the only goal here is not, as is too often naively assumed by researchers in artificial intelligence, (or by others who consider the sort of reflection involved in examining such arguments a waste of time) merely to refute the arguments in question. Of course, I would like to believe that my logical analyses have been correct and that I have, indeed, succeeded in refuting these arguments, but this is no more than half the story. The real lasting value from these investigations and the time spent on them is to be gained in what we are able to learn from the process. As with Plato's famous dialogues, it is almost more important what ground we cover than what discernible final destination we approach. In attempting to reply to either side of the argument, we are forced to examine the foundations of our knowledge and suppositions about the nature of intelligence and cognition.

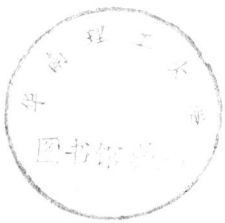
As we explore the territory, we find gaps in that knowledge and errors in those suppositions. These in turn not only provide us with the means of evaluating reports of progress in artificial intelligence research, but also call to our attention some of the very topics and issues which should be investigated in order to make significant progress in that research. Perhaps the clearest example of this can be seen in the work of Hubert Dreyfus. While his case is not made for an argument *in principle* (i.e., one which shows the *impossibility* of artificial intelligence), if we interpret his remarks instead as a criticism of work to date in artifi-

cial intelligence, we find that not only are Dreyfus' criticisms accurate, but that they suggest the directions and issues that should be explored by artificial intelligence workers in future research.

ACKNOWLEDGMENTS

Thanks are due to the following: Henry Kyburg for his criticisms and advice; Richard Taylor for his understanding and philosophical open-mindedness; Ken Sloan, Robert Holmes, Madhab Mitra, Don Perlis, and Skip Mihalyi for various conversations on machine intelligence; Cynthia Helms for her criticisms of an earlier version of Chapter 3; Judy Dering for listening to parts of Chapter 5; and Lewis W. Beck for his continual encouragement and kindness.

Acknowledgment is also due to Alan Turing, Keith Gunderson, Hubert Dreyfus, Richard Taylor, and Norman Malcom, without whose writings the present work would have been literally impossible.



CONTENTS

MACHINES AND INTELLIGENCE

1
THE TURING TEST

In his article "Computing Machinery and Intelligence," after a brief remark to the effect that he will not attempt to define "thinking," Alan Turing directs our attention to another problem which he says is closely related to the question "Can Machines Think?" but differs from it by being expressed in "relatively unambiguous words." The new problem is to be posed in terms of the imitation game:

> It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A.' The interrogator is allowed to put questions to A and B. . .[1]

A's object is to fool the interrogator, while B tries to aid the interrogator in discovering the identities of the other two players. For the game to present any challenge, of course, the players A and B must not be visible to the interrogator. Turing suggests that they be in a

separate room, and communicate with the interrogator only by means of a teleprinter (teletype). Turing poses his new problem:

> We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?'[2]

This new question (which we will refer to henceforth as the Turing Test) is not Turing's only contribution to the debate about thinking machines. In the same article, "Computing Machinery and Intelligence," he also discusses the general nature of computing machines and examines numerous arguments against machine intelligence. In other parts of this work we will be looking at some of the same arguments that Turing discusses, and we may therefore find occasions to refer back to Turing's discussions as they become relevant. In this section, however, we will be concerned with the Turing Test and the question of its validity. We will begin by considering the arguments Turing advances in support of his test.

Turing remarks that the condition which prevents the interrogator from seeing, touching, or hearing the competitors helps draw a sharp line "between the physical and intellectual capacities of a man." This is advantageous because:

> We do not wish to penalize the machine for its inability to shine in beauty competitions, nor to penalize a man for losing in a race against an airplane. The conditions of our game make these disabilities irrelevant. The 'witnesses' can brag, if they consider it advisable, as much as they please about their charms, strength or heroism, but the interrogator cannot demand practical demonstrations.[3]

This seems fair enough; after all, we are concerned with abilities of the machine to think, not with its physical appearance, nor its strength, and so on.

In addition Turing claims that "The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavor that we wish to include." Turing gives some examples:

> Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess?

A: Yes.

Q: I have my K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: (After a pause of 15 seconds) R–R8 mate.[4]

While they are admittedly points in the test's favor, these two aspects of the test hardly seem to give us conclusive reasons for replacing the question "Can machines think?" by the question "Can machines pass the Turing Test?" For whatever means we might have used in answering the first of these questions, it seems clear that we need not have been concerned with such things as the physical appearance or strength of the machine. It also seems clear that we were free to examine the machine's abilities in widely divergent areas. Thus neither of the two points mentioned give us really good reasons for adopting the new question. I will now suggest one point in favor of the test that Turing does not mention explicitly, but which I think provides at least some of the motivation for Turing's proposal.

One of the most important motivations for Turing's proposal seems to be the fact that the test guarantees completely equal treatment of the man and the machine. Thus we are not allowed to ask (as Scriven suggests in his paper, "The Compleat Robot") whether we are entitled to say that the machine thinks *despite* the fact that it is not a man. Scriven says:

> The performatory problem here is whether a computer can produce results which when translated, provide what could count as an

> original solution or proof *if it came from a man.* The personality problem is whether we are entitled to call such a result a solution or proof, despite the fact that it did *not* come from a man. . . . If it transpires that there are no essential performatory differences at all, we shall then consider whether we are entitled to apply the terms in their full sense.[5]

If we cannot tell the man and the machine apart (since we know them only as "A" and "B" and can't see them) then we cannot discriminate against the machine; we are forced to use the same sorts of criteria for both. We cannot say "*This* one is the machine, let's be extra cautious with *it*." All we can do is to apply whatever criteria we can come up with in order to arrive at an educated guess as to which is the man and which is the machine. The manner of testing the man and the machine is thus the *same*, namely the Turing Test.

That this is very likely part of the real motivation behind Turing's proposal can be seen from his reply to what he calls "The Argument From Consciousness." There he gives us some more arguments in support of his test; and since the importance of that portion of Turing's article seems to have been ignored by both his supporters and his critics, this provides us with an additional reason for reviewing his discussion of that argument.