O231
G347

8060938
5

PHILOSOPHICAL FOUNDATIONS OF CYBERNETICS

F. H. GEORGE, M.A., PhD.
Director, Institute of Cybernetics
Brunel University, United Kingdom

华南工学院
外国教材
中心图书室藏书

E8050938

ABACUS
PRESS

First published in 1979 by

ABACUS PRESS

Abacus House, Speldhurst Road, Tunbridge Wells, Kent TN4 0HU

© Abacus Press 1979

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Abacus Press.

British Library Cataloguing in Publication Data
George, Frank Honywill
Philosophical foundations of cybernetics
(Cybernetics and systems series).
1. Cybernetics – Philosophy
I. Title II. Series
001.53'01 Q310
ISBN 0-85626-163-7

Printed in Great Britain by
W & J Mackay Ltd. Chatham, Kent

1. PHILOSOPHICAL FOUNDATIONS OF CYBERNETICS

CYBERNETICS AND SYSTEMS SERIES

Editor in Chief

J. ROSE

Director General,
World Organisation of General Systems and Cybernetics

Other Titles in Preparation

2. Fuzzy Systems
3. Artificial Intelligence
4. Management Cybernetics
5. Economic Cybernetics
6. Cybernetics and Society
7. Models and Modelling Systems
8. Automation and Cybernetics
9. Medical Cybernetics
10. General Systems Theory
11. Computers and Cybernetics
12. Neurocybernetics

此为试读，需要完整PDF请访问：www.ertongbook.com

SERIES PREFACE

The inter- and trans-disciplinary sciences of cybernetics and systems have made tremendous advances in the last two decades. Hundreds of books have been published dealing with various aspects of these sciences. In addition, a variety of specialist journals and voluminous post-conference reports have appeared, and learned societies, national and international, established. These substantial advances reflect the course of the Second Industrial Revolution, otherwise known as the Cybernetics Revolution.

In order to extend the readership from experts to the public at large and acquaint readers with up-to-date advances in these sciences which are rapidly achieving tremendous importance and which impinge upon many aspects of our life and society, it was considered essential to produce a series of concise and readable monographs, each concerned with one particular aspect. The twelve topics constituting the first series are as follows (in alphabetical order).

Artificial Intelligence, Automation and Cybernetics, Computers and Cybernetics, Cybernetics and Society, Economic Cybernetics, Fuzzy Systems, General Systems Theory, Management Cybernetics, Medical Cybernetics, Models and Modelling of Systems, Neurocybernetics, Philosophical Foundation of Cybernetics.

The authors are experts in their particular fields and of great repute. The emphasis is on intelligible presentation without excess mathematics and abstract matter. It is hoped each monograph will become standard reading matter at academic institutions and also be of interest to the general public. In this age of enormous scientific advances and of uncertainty concerning the welfare of societies and the very future of mankind, it is vital to obtain a sound insight into the issues involved, to help us to understand the present and face the future with greater confidence.

J. ROSE
Blackburn

PREFACE

In writing this book I have been guided by a number of considerations. The first and most important is to supply a clearcut picture of what the central themes of cybernetics entail. If we state baldly that "machines could be made to think", then we have to examine what it is that we mean exactly, and what such a statement entails. We have to examine evidence and above all be clear about the philosophical implications. In this regard I have been greatly influenced by the late Dr. Alan Turing who set out in his article "Computing Machines and Intelligence" (published in 1950) the essential issues. There he considered some of the likely objections to such a view and I have followed these up and have gone on to try to make the whole matter more explicit.

The second consideration was in terms of what I had previously said myself on the subject. I hope that what I now say is consistent with such previous statements, even if I now place the emphasis somewhat differently.

A particular aspect of this second consideration is that I wish to discuss cybernetics in philosophical terms, but with emphasis on the central themes of cybernetics. This was read as a restriction on the possibility of here examining epistemology, logic, truth, meaning and the like in great detail – perhaps even from a cybernetic point of view. But I have tackled this second more philosophical matter in a separate book which should appear soon after this one. The reason for mentioning this matter is to ensure that the more philosophically inclined reader should not feel disappointed at the relatively small space given to the more traditional philosophical problems.

Certain of these traditional problems, such as the mind–body problem, are so central to our purpose that they are discussed, but others such as that of ontology receive only brief mention.

I would like to say also that I believe the relation between cybernetics and philosophy is not only close, but it is a two-way relationship, since I believe that one's philosophical views are clarified by cybernetic thinking. Similarly any possibility of being clear about cybernetics without careful philosophical thought is inconceivable.

This last statement represents my view that science and philosophy are very much more closely related than one would guess from a casual reading of most books or journals on either subject. The obvious exception is that of the work on the philosophy of science, but even this does not deal with the relationship I am primarily concerned with. There is a deeper relationship between basic philosophical thinking and science – rather especially cybernetics – that I am hoping this book will bring out. It is not just that philosophy should be invoked to analyse scientific activities, but that science should be used to analyse philosophical activities.

In looking at cybernetics in this manner, I owe a number of acknowledgements to different people. In the first place I have submitted, over the years, a number of postgraduates at Brunel to my views and have, without doubt, clarified my own as a result of the feedback derived from their views.

In particular I owe a debt of gratitude to Mr. W. J. Chandler, Director of Corporate Planning at Reed International, for a whole series of discussions which surround his ideas on the science of history and planning, and of which I have openly taken advantage.

I also owe a special debt to Mr. L. Johnson of the Department of Cybernetics at Brunel for his reading and constructive comments made on this text. Also my wife and my secretary at the University, Mrs. P. M. Kilbride, my elder daughter Mrs. C. E. Smith and my younger daughter Miss Karen George for varying degrees of help, both direct and indirect.

As usual one had to add that the extent to which the final result may seem adequate is entirely my responsibility.

Frank George
Beaconsfield

8060938

华南工学院
外国教材
中心图书室藏书

CONTENTS

8060938
5

华南工学院
外国教材
中心图书室藏书

Chapter 1

ARTIFICIAL INTELLIGENCE AND THE INTERROGATION GAME

This book is concerned with the philosophical background of cybernetics. As a science, cybernetics is the science of communication in animals, men and machines, and we shall not in this book seek to justify it *as a science*. This is because we feel that it needs no justification other than the fact that it exists and is, in our view, satisfactorily progressing along a fairly well-defined line. As a science, it can be judged by its practical pay-off and from this point of view it is not necessarily of great importance whether we taken one philosophical view of it rather than another.

We shall, however, be discussing cybernetics from a philosophical point of view. One of the basic questions we shall be considering is "whether or not machines can be made to think?" The philosophical importance of this is obvious, but the scientific importance is very much less obvious. It does not really matter, provided the science is supplying a good 'spin-off', whether or not in the end it can fulfil (or wholly fulfil) this goal. So it is, that to this extent the science of cybernetics, which is concerned with artificial intelligence and its application in all sorts of other fields, such as behaviour, biology, economics, business, education, etc. is not the subject of our discussion.

The point we have already made about "machines thinking" as being one of the central points of cybernetics, will also be the central theme of our own discussion. It will be stated in the form[1] "*Could* machines be made to think?" Sometimes this has been phrased in the form "*can* machines be made to think?"[2] but we are not concerned with the relatively unimportant sense in which this has already been proved to be possible, when compared with the far more important sense in which we think it could be made possible. It is best from our point of view to throw the question into the future and say, regardless of whether or not it is possible now, is it possible in principle?

We should be clear from the start that a question put this way is easily misunderstood, if for no other reason than various key words have a variety of different possible meanings. In other words, if we defined the word 'machine' and defined the word 'think' then we could, without too much trouble settle our question in the affirmative or negative, as a direct result of our choice of definitions. We could say that machines by the very nature of things (i.e. by our definitions) are precisely those systems which do not think, that are automatic and unthinking, therefore to ask whether or not they think is an absurd question. We could, on the other hand, take a very much more general definition of the 'machine' to include human beings for example, in which case to ask whether or not they think is equally absurd because they obviously do, since they now include a class of systems which manifestly (by definition) thinks. Therefore we have to consider other ways of phrasing our central theme which makes it more intelligible, and easier to handle.

The first difficult word to define is undoubtedly the word 'machine'. We really want to talk about systems "capable of being manufactured in the laboratory". The artificial manufacture of the system is the important thing; not whether it is machine-like in any other sense. We are not thinking of a machine such as a potato peeler or a motor bicycle; we are thinking of an artificially constructable system which is capable of being manufactured in a laboratory and which also has the properties of adaptability which characterise human beings. We have to be careful here to distinguish between artificial insemination, for example of a human, which provides another human, and the laboratory manufacture of a seed which is capable of growing in our own artificially prepared environment and becoming human-like.

There is also a further difficulty because when we ask about the possibility of machines (in the complex sense of artificially constructable systems) thinking, we do not necessarily mean in a human-like way. However, we are bound to use the human being as a yardstick, and ask whether or not we could manufacture a system which is capable of thinking with the same degree of efficiency as a human being. It seems, however, likely to follow from this, provided we can understand the general principles by which human beings can think as effectively as they do, and then reproduce these principles in an

此为试读,需要完整PDF请访问: www.ertongbook.com

artificial system. Given this situation, the possibility of producing a system which is superior to man in its abilities is fairly straightforward. Not that we wish to make the claim at this point that we can make machines that can think more efficiently than humans; our argument would rest sufficiently on making machines that can think at least as efficiently as humans.

The second main word 'think' provides another obvious difficulty, since some people use this word to apply purely to what humans do and indeed not only to what humans do, but what they are conscious of doing. We would want to say that thinking is a process of manipulating symbolic representations of events, and the process of learning and adapting as a result of these manipulations, as well as solving problems and formulating plans etc., without necessarily being conscious of the process one is going through, and without necessarily being a human. This sort of definition is behaviouristic by inclination, and does not insist as some people do (more often than not philosophers talking of thinking) that thinking is a process necessarily involving consciousness. This, of course, is not to say that much of what we call thinking is not actually a conscious process, but that is another question.

THE INTERROGATION GAME

We will have made the point clear that talking about the possibility of machines thinking implies something fairly special in the meaning of the word 'machine'. It also implies something fairly clear-cut by way of the meaning of 'thinking' where we mean to use the human being as a yardstick. It leaves the matter open as to whether the possibility of synthesising (as opposed to simulating) a human being could actually use methods for effective thinking and problem solving – other than human methods. We do not need though to discuss that particular question at this point. However, we do want to look at the question of the Turing interrogation game to be quite clear that the system we produce is not necessarily human-like in its construction.

There is a parlour game that has sometimes been played (though not in the experience of the present writer) whereby you try to decide, as a result of asking questions, whether a human being is a man or woman. Clearly this is subject to the constraint of not being able to look at the person in question. So you try to formulate questions which would elicit answers that should somehow serve to give the questioner a clear picture of to which of the two sexes he is talking.

Turing has suggested an adaptation of this interrogation game in order to distinguish the human from a machine; he believed that this would be an effective way of defining the concept of a machine's ability to think. If you can carry out an interrogation game with a human being compared with an artificially constructed system (as opposed to another human being) and if you find it impossible to tell which is which, then you have to accept the fact that a machine can think as well as a human being. In using the phrase 'as well as' we are not concerned so much with precise relative abilities in every sphere, only that in general the quality of human-like thought is equally attributable to both.

It could be objected that this does not compare with the human-like quality of thought, as much as the human-like responsiveness of the two systems. The answer to this is that the problem of 'other minds' arises just as much when comparing a human being with oneself. It is not possible to tell whether other human beings think, all one can tell is that they behave (or do not as the case may be) *as if* they thought. So we shall have to settle for this criterion when comparing a machine and a human being; we shall have to decide whether it behaves *as if* it thought.

The importance of the interrogation game lies in the fact that it is saying, in effect, that the artificially constructed system could be a digital computer, albeit a fifth or sixth generation computer. It could on the other hand be an electronic system of some kind and certainly does not need to be made, as a human is, of colloidal protoplasm. All that matters is that it should behave in a similar way in a similar sort of situation.

OBJECTIONS TO THE MAIN THEME

Our main theme is that *machines could be made to think*. We are answering the question posed by our main theme at least to the extent of saying that we can think of no reason to doubt the possibility. We now start to cast around for possible objections to our viewpoint. Some of these objections will receive the most detailed analysis in separate chapters which follow; others will merely without detailed analysis be mentioned.

The first objection, which we shall not treat in great detail, is the theological objection. This says in effect that thinking is a function of man's immortal soul, and must be something attributable to man and man alone, and that no other system could possibly achieve it. This in some part is like the argument which says thinking is human-like and human-like only (though not necessarily on theological grounds) and therefore cannot apply to machines. We shall simply assume the wrongness of this argument and leave a discussion of a theological kind outside the text altogether. In saying this we should perhaps just mention that we have here the support of D.M. Mackay, who, while having strong positive theological views, would still accept the fact that machines could be made to think in the sense that we intend it to be understood in the text.

Next, there is what Turing calls the 'heads in the sand' view which simply says that it would be 'quite dreadful' if machines could be made to think, and so as a result rival human beings. In much the same way as it would be dreadful if we found another species overtaking us in our ability. On purely biological grounds there seems no reason to doubt the possibility of another species overtaking the human species, and by the same sort of argument it seems pointless merely to say it would be 'quite dreadful' and feel that this is the counter-argument, so we shall also altogether neglect this type of counter-argument.

The third type of argument we shall consider in Chapter 3. This is an argument from the point of view of the foundations of mathematics. Basically, it is an argument based on Gödel's theorems[3] and related theorems due to Turing[4], Church[5], Post[6] and others. The essential features of these arguments are based on the fact that you seem not able to construct an axiomatic system in which both its own completeness and its own consistency can be demonstrated from within the system. In other words, there are certain statements or features which we would accept as necessarily being within the system which we cannot demonstrate in our axiomatic system when that axiomatic system is used to investigate its own characteristics.

We shall try to show in the next chapter but one that the Gödel arguments as they are sometimes called do not really serve as a barrier to our main theme. However, this is such a complicated matter that we will not attempt to provide the counter-argument here.

Another type of argument used as a contrary to our main theme is the argument of consciousness. We shall be treating this from various points of view, since we want to discuss the relation of consciousness as it seems to occur in humans, with the possibility of an equivalent state occurring in machines. We also wish to discuss the problem of free will, and the problem of creative ability. Various aspects of consciousness can be broken down into various possible counter-arguments and these will be examined in considerable detail. We shall say no more about it at this point but merely notice that the property of consciousness, which seems to be a characteristic of human beings, will be seen by some as a barrier to the possibility of making machines capable of performing the same sort of activities as human beings, on the grounds that they (the machines) could not possibly have consciousness. We believe this is a false argument too and we hope to show it, but not only in one way, rather in various ways.

Another argument considered by Turing as a counter-argument is what he calls the argument of 'various disabilities'. This in essence, says that there are various things which you cannot make a machine do that a human being can do. One of these 'things' is that the bodily structure of a human being would be extremely difficult to produce by any mechanical engineer however sophisticated. We shall not argue this particular point because we think that the intelligence shown, which is the basis of our argument, is independent of the structure which shows that intelligence, and therefore we do not necessarily want to produce a system made in the same way as a human. There are some doubts about this view since some people feel that the fabric of manufacture is closely bound up with the system's performance. We will bear this objection in mind.

Other disabilities which should be mentioned are the inability to reproduce oneself and the inability to function under conditions of error. Von Neumann[7] has shown that both these arguments can be overcome, and that artificially constructed systems can reproduce themselves, and furthermore however much error there may be in the functioning of the system, provided that the system is sufficiently complex, it will survive that error and correct it if necessary.

Turing has actually used the argument that if you specify precisely what it is that *cannot* be done by our artificially constructed system then from the description of what it is that it cannot do we will manufacture the system to do it. This argument, while persuasive in part, is not necessarily wholly acceptable to the human-like claims of an artificially constructed system and where we say that it cannot do a certain thing. We may only be able to point to the end result, however, without saying how it is achieved, and therefore, not necessarily give enough information to make it possible to reproduce what it is that is required or is missing, and is claimed to be impossible to reproduce. We shall be looking at this question of various disabilities throughout the text as it rears its head in various different contexts.

There is one further argument which we will consider and that is that a computer, or any other artificially constructed system, only does what it is made to do by its programmer. This is a view that was held by Lady Lovelace. It is a popular fallacy that computers can only do what the programmers make the computers do. The fallacy arises from various considerations. One is the failure to remember that human beings only do what they are programmed to do, although they are programmed by various different features of the environment, including parents, teachers, etc. and are adaptable and change according to changing circumstances. Now in this sense it is perfectly true to say that computers can only do what they are programmed to do, but they can certainly be given exactly the same flexibility as humans. In other words, various people can program them and they can be made adaptive so that they change and function in changing circumstances.

In other words, if you are making an intelligent machine to play chess, for example, then to make it merely reproduce a set of standard moves which had been thought out by the programmer would be perfectly useless. It is absolutely essential that it should be given only starting programs all of which it is capable of changing in the light of its particular experience. We can, therefore, firmly disregard as an important objection to our main theme the argument that a computer only does what it is programmed to do, it has no relevance whatever. Nevertheless we shall be repeating this point more than once in the course of this book, since it is such a widespread misunderstanding that it needs to be emphasised frequently that it is *just* a misunderstanding.

Turing's interrogation game and the objections it gives rise to automatically involve us in some philosophical and scientific discourse. This in turn means that we are bound to be involved in the philosophy of science, as well as epistemology, ontology and the like.

The philosophical implications of cybernetics do take us straight into a number of philosophical issues which could be regarded these days as virtually 'off the shelf'. They are the so-called 'mind–body problem', the problem of 'other minds', 'free will' and questions, notoriously difficult to deal with, such as that of 'consciousness'.

To some extent these issues thread through our whole book and are fundamental to the philosophy of cybernetics. If, for example, you argue that minds are something private and characteristic of human beings and human beings alone, then, *by definition*, the question of making 'machine minds' is a contradiction. We must therefore be careful not to become involved in linguistic absurdities or obscurities if they can be avoided. Bearing in mind, the problem as seen from the viewpoint of the interrogation game (this particularly deals with 'other minds'), let me say that the questions of 'free will' and 'consciousness' are later discussed in detail. So now to complete this introductory chapter, we will explicitly say a few words on the 'mind–body' problem and how it can be regarded from the viewpoint of the cybernetician.

There are many ways of looking at the mind–body problem, but we will start by considering what are sometimes called cognitive terms. Sommerhoff[8] has recently made the point which is often made in some form or another:

> . . . . the deplorable mistake of Behaviourists of interpreting the meaning of all mental state concepts as synonyms with the respective behaviour dispositions.

These behaviour dispositions are not, argues Sommerhoff, what we commonly mean by such cognitive terms as 'perception', 'learning', etc. He further argues that Ryle[9] was influential in the process of interpreting mental state concepts (words) as dispositions so that the shift from private descriptions to public descriptions occurs, and by implication misleads. Sommerhoff then says that Ryle believes that this (behavioural disposition) is what we mean all the time by these mentalistic terms and that, says Sommerhoff emphatically, is not correct.

Regardless of what Ryle thought about the matter, let us take this as our starting point, as at least being a more plausible argument on behalf of what we might call 'behaviourism' than that advanced by more extreme supporters of such a view: for example, Watson.[10].

There is nearly always a problem of translating from terminology

此为试读，需要完整PDF请访问：www.ertongbook.com