ROUTLEDGE REVIVALS

Goals, No-Goals and Own Goals

A Debate on Goal-Directed and Intentional Behaviour

Edited by
Alan Montefiore and Denis Noble

Routledge Revivals

Goals, No-Goals and Own Goals

First published in 1989, *Goals, No-Goals and Own Goals* presents a stimulating debate between three scientists and three philosophers about the significance and nature of goal-directed and intentional behaviour. At one extreme David McFarland brings into radical question the need for either of these concepts, at least in the scientific study of animal behaviour. At the other extreme Alan Montefiore argues that such concepts are indispensable to any explication of the meaningful use of language and that we must therefore acknowledge their importance in understanding the nature of human behaviour. Denis Noble uses arguments drawn from computer science and physiology to show that it is incorrect to regard intentions as causes of neural events, even though it is correct to regard intentionality as responsible for our actions. Shawn Lockery outlines how intentional behaviour might be subjected to physiological study. Kathy Wilkes widens the debate by asking some basic questions about the nature of explanation and finally, Daniel Dennett argues how the study of animal behaviour might inform research in Artificial Intelligence. This book will be a useful resource for scholars and researchers of cognitive science, philosophy, psychology, linguistics and physiology.

Goals, No-Goals and Own Goals

A Debate on Goal-Directed and Intentional Behaviour

Edited by Alan Montefiore and Denis Noble

Routledge
Taylor & Francis Group

First published in 1989
by Unwin Hyman Ltd

This edition first published in 2021 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN
and by Routledge
605 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

A Library of Congress record exists under LCCN: 88033932

ISBN 13: 978-1-032-02862-0(hbk)
ISBN 13: 978-1-003-18555-0(ebk)
ISBN 13: 978-1-032-02864-4(pbk)

GOALS, NO-GOALS AND OWN GOALS

A debate on goal-directed and intentional behaviour

Edited by

Alan Montefiore & Denis Noble

Balliol College, Oxford

LONDON
UNWIN HYMAN
BOSTON SYDNEY WELLINGTON

This volume was prepared, proofed and passed for press
at the University of Oxford.

Published by the Academic Division of

Unwin Hyman Ltd
15/17 Broadwick Street, London W1V 1FP, UK

Unwin Hyman Inc.,
8 Winchester Place, Winchester, Mass. 01890, USA

Allen & Unwin (Australia) Ltd,
8 Napier Street, North Sydney, NSW 2060, Australia

Allen & Unwin (New Zealand) Ltd in association with
the Port Nicholson Press Ltd, Compusales Building, 75 Ghuznee Street,
Wellington 1, New Zealand

First published in 1989

British Library Cataloguing in Publication Data
Montefiore, Alan
Goals, no goals, and own goals: a debate on goal-directed
and intentional behaviour.
1. Intention, — Philosophical perspectives
I. Title II. Noble, D. (Denis)
128'.4

ISBN 0–04–445341–8

Library of Congress Cataloging-in-Publication Data

Goals, no-goals, and own goals

Bibliography: p.
Includes index.
1. Action theory. 2. Agent (Philosophy). 3. Intentionality (Philosophy)
I. Montefiore, Alan. II. Noble, Denis.
B105.A35G63 1989 128'.4 88–33932
ISBN 0–04–445341–8 (alk. paper)

CONTENTS

PART III THE POSITIONS DEBATED

PART IV A CHALLENGE RENEWED

PREFACE

We hope this book will be of interest to a very wide range of readers, many of them students in philosophy, psychology, computing and in the behavioural and physiological sciences. But even those of our readers who may not think of themselves as students may nevertheless find it helpful to know what to expect in this book and how it might best be read.

The book is organised into four parts. Part I consists of three introductory chapters. There are three chapters here, because we are aware that readers with very different backgrounds will require different kinds of introduction. The first explains how the book came to be written and what are its purposes and organisation. The second provides a brief account of some of the relevant philosophy, while the third acts as a more scientific introduction. We hope these chapters will make the book more accessible to those readers who require some introduction to our debates, but we are aware that much of the material here may be unnecessary for many others. Those who are already familiar with the kind of interdisciplinary debate to be found here, may find it possible to embark straight away on Part II where the positions to be debated are laid out. In Part III the initial positions outlined in Part II are discussed, and finally in a single chapter, in Part IV, one of the authors presents his reactions to the debate.

Those who embark directly on Parts II–IV may later wish to read some of Part I as a concluding overview of the book.

Certainly, parts of these introductory chapters may only be fully appreciated *after* having engaged in the debate itself. That also means that, read as introductions, some of the points made there will have to be taken 'on trust', waiting for the real debate to begin to see what we are trying to say.

A word about bibliographies. Wherever it has seemed that it might be appropriate and helpful, the contributors have added a few suggestions for further reading at the ends of their own chapters with a note of explanation as to what might be expected of the books suggested. At the end of the book, we have appended a much longer list of all those works to which reference has been made in the main text; in a separate section we have also added a short list of works to which reference has not been made, but which might prove of further interest to readers wishing to pursue the discussion in one direction or another. Many of the works mentioned under one or other of these heads themselves contain helpful and sometimes extensive bibliographies and/or suggestions for further reading on specific topics.

A book of this kind is almost impossible to index since the items that might be indexed are diffused throughout the book. Each index item would have required reference to a very large number of pages. We have instead indicated in the bibliography the pages on which each reference is cited. This should be useful to those who wish to find where a particular paper or book is discussed.

Inevitably the debates do not follow a single straight line. It may sometimes be found necessary to weave back and forth between the chapters in Parts II & III to get the full flavour of the discussion. The significance of some of the arguments will only become fully clear when the other chapters have been read — or perhaps re-read — in the light of a later argument. If you find you need to do this, then you will be retracing the spirit of the original university seminars and the many associated less formal discussions that led to this book being written.

Alan Montefiore and Denis Noble

Acknowledgements

We should like to thank many of our colleagues for their helpful and critical comments at various stages in the work for this book, and in particular Dr Susan Greenfield and the anonymous publishers' readers for their valuable suggestions, nearly all of which have been acted on in revising the book for publication.

The Wellcome Trust has been helpful in two ways. First, by providing Shawn Lockery with a research grant during his work in Oxford and, second, through a Major Equipment Grant for the SUN computer system on which the text was typeset using TEX and Textcode. (TEX is a trademark of the American Mathematical Society, and Textcode is a trademark of Oxsoft Ltd.)

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

PART I
INTRODUCTION

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

CHAPTER 1

GENERAL INTRODUCTION

Alan Montefiore and Denis Noble

This volume is intended to provide an interdisciplinary approach to some of the many intertangled problems connected with the identification, characterisation, understanding and explanation of goal-directed and intentional behaviour. In fact, this very opening sentence of what is 'intended' as our own introduction to our own volume also presents itself as a telling example of the problems with which we are concerned; for it contains what appears as an unabashedly straightforward declaration of would-be intention. But one of the most intricate set of issues addressed in the discussions that follow turns around the disputed significance (or lack of significance) of the fact (if it is indeed a fact) that we human beings, we speakers and writers of language, seem to be unable to argue or even to think ourselves free of such reliance on reference to or signalling of our own intentions.

This volume has, then, six authors, half of them professional scientists, half of them professional philosophers. Their contributions are not, however, to be found aligned on opposite sides of some imaginary line separating the 'scientific' from the 'philosophical' point of view. On the contrary, on just about every major issue the scientists were to be found tending towards disagreement with other scientists just as the philosophers with other philosophers, while members of each 'professional group' could look to find allies as well as dissentients from among members of the other. The lines of debate cross and re-cross that of any boundary that might be drawn to distinguish scientists from philosophers. That boundary is here more discernible in the differing experience and expertise of those who have been trained to work in either one field or the

other. But even this boundary contains numerous crossing points. The scientists here involved are well read in at least some major areas of the professional philosophical literature = indeed, one of them has contributed to it; while the philosophers have been particularly concerned, so far as they are competent to do so, to take the findings of the relevant scientific disciplines into account.

However, this situation is not, it should be stressed, accidental; the interdisciplinary nature of this volume does not consist in its simply having been put together on the basis of *ad hoc* contributions specially invited from representatives of different disciplines. Rather it has its origins in successive series of seminars, together with many surrounding discussions, in which all but one of the present authors have been engaged in Oxford over the last two decades. This is not to say that any one of those five has taken part in all of those seminars and discussions. Still less should it be taken as either forgetful or unappreciative of all those many other participants who have appeared in these ongoing debates = colleagues and students alike, philosophers, psychologists, physiologists, animal behaviourists and many others = whose criticisms, objections, questions, suggestions and general stimulus have so much contributed to their enrichment. Moreover, not only is our sixth contributor, Daniel Dennett, well known to all the rest of us through his writings on our common topics; to a number of us he is known through personal encounter as well. In short, this book has grown out of a whole series of discussions between a number of people of very different backgrounds and experience but with common and overlapping concerns, and in particular the common conviction that none of those concerns has any sensible likelihood of finding satisfactory pursuit other than in such cross-disciplinary co-operation and debate. It goes without saying that the end of this book is in no way the end of our discussions. It is rather, we hope, the occasion for others to join in.

The discussions that form the background to this book have, then, been going on for what is by now a long time. For a number of the present contributors they had their origins in our several reactions to Charles Taylor's *The Explanation of Behaviour* (Routledge and Kegan Paul, 1964). The earlier seminars led to publication in various forms, including the *Analysis* debates in 1967 between Noble and Taylor, the Aristotelian Society symposium of 1971 on 'Final Causes' between Timothy Sprigge and Montefiore and parts of Anthony Kenny's *The Five Ways* (Routledge and Kegan Paul, 1969). The present book, though clearly influenced

by those earlier debates, takes a largely new look at the issues, being more directly based on a renewed set of seminars (and some intensive related discussions) held in Oxford over the last three or four years.

These seminars provided an extended opportunity in which to try out ideas in a context of critical discussion. However, it did not seem sensible to try to recreate in this book the atmosphere of the seminars themselves by way of some sort of reconstructed transcript — even supposing that we might have been able to present a plausible reconstruction. We decided, therefore, to start more or less from the point that the seminars had reached. Each author was asked to write an opening chapter presenting the position that he took himself to have reached at that stage and the problems with which he was most immediately preoccupied. Once these chapters had been circulated and discussed in draft form, second chapters were written in which each had an opportunity to react to what had been produced by the others or to elaborate further on any points of his or her own which he or she wished to develop. Those draft first chapters have here been modified only to the extent of seeking to eliminate sources of distraction or unfruitful misunderstanding and to improve their presentation. In general, even though some of us may subsequently have been led to modify or even to retract some of our first chapter views or formulations, they have nevertheless been allowed to stand here in the interest of preserving the onward movement of debate. For, as will be evident enough, much of what has been written in the second chapters takes the form that it does as direct response to what had been written in the first.

What of the order in which the different contributions have been placed? It would be a mistake to attribute too much importance to it. None of us set out to write either his or her first or second contribution with any particular order already in mind, and the order on which we have finally settled was established entirely *ex post facto* in primary response to the evident necessity of having some order or another. Nevertheless, we do see real significance in our decision to place David McFarland's opening statement first, because, as a strikingly bold and articulate expression of a set of views that, in some largely unexpressed form or another, are probably taken for granted by a wide variety of working scientists, it represents a standpoint that did in fact serve as a main organising or focusing principle for many of our more recent discussions, and that still so serves in many of the discussions that are to be found

in this book. Alan Montefiore's opening statement comes second because it represents, on many of the central issues at any rate, a diametrically opposed view. From there on we have simply continued by way of an alternation between scientists and philosophers that conveys very fairly the thoroughly interdisciplinary nature of our debates.

The same broad considerations apply to the ordering of the second round contributions. Its inevitable overall linearity may be somewhat misleading, however, in as much as these second round contributions are not simply and straightforwardly responses to the whole set of opening statements, but, in the case of those participants resident in Oxford at any rate, reflect also the fact that discussion among them has naturally proceeded as they have continued to work on these matters. Indeed, the best way to read these second round contributions is as both the record and the continuation of a multilateral debate, with all the internal cross-currents that any such discussions are bound to generate. If anyone should ask why any particular second round contribution takes up the particular issues that it does, while failing to take up others which, no doubt, it might equally well have taken up, the answer is again to be found in the nature of discussion itself; one inevitably tends to respond to what seems most immediately pertinent or challenging in the light of one's own immediate preoccupations. (It may also be, of course, that points are sometimes not taken up because the author concerned — wrongly perhaps — takes his agreement with them for granted, and for granted also that, in the light of his general position, his agreement must here be transparent to all.) All of us — indeed, this has been one of our main editorial troubles — keep on having further thoughts on old thoughts, or thinking of new things to say in further reaction to what has been said by one or another of our fellows. But this too is of the nature of an on-going discussion; there would be little point in inviting others to join in if one did not know it to be essentially incomplete.

Our final section differs from the first two in that it consists of one third round contribution alone. The reason for this is not — it need hardly be said — that David McFarland is the only one among us to feel the urge to return to the argument, to take up the points that have been urged against him and, in so doing, to carry forward the working out of his own position. But just as we have placed his first contribution first in recognition of the pivotal role which his (as some of us have thought, rather extreme) views have

played in the development of the rest of our discussions, so it has seemed appropriate to ask him to conclude the volume with what is, very clearly, not a summing-up sort of conclusion, but rather the opening thrust of the next round of debate. For, to repeat, the topics of this debate are urgent, the discussion remains very much open and if there are here no third round contributions from the other five participants concerned, that is certainly not because they would have nothing to say in further reply.

While the results of all this, and in particular, of course, the opening statements as they here stand, reflect many of the seminar discussions, they also differ from them in certain quite substantial ways. These differences lie primarily in the fact that by the time that the contributors got down to preparing their opening statements many of the arguments that they had earlier been pursuing in the seminars themselves had virtually disappeared. These were arguments that those concerned felt to have been settled, or to have been shown to be of no substantial importance. Nevertheless, the issues in question include some on which it is important to comment here in order to help the reader pick up the threads at the point at which the published debate takes off.

A major issue of the earlier seminars, and one that remains on the surface of even much later debate, was whether the choice between teleological and non-teleological forms of explanation for the occurrence of behaviour that might in principle be specified in descriptively neutral terms, could be decided on straightforwardly empirical grounds. This question figured very largely in the debate on Taylor's *The Explanation of Behaviour*, and echoes of that debate are still to be found in Noble's opening chapter (chapter 6). It no longer, however, figures as a key issue; for there now seems to be common agreement that the distinction between the conceptual and the empirical cannot usefully or plausibly be made hard and fast in any general or overall way. Thus, in trying to develop criteria for the identification and characterisation of goal-directive behaviour, and in analysing the peculiar nature of intentional behaviour in particular, the empirical–conceptual dichotomy no longer seems to be of central importance. Still, it would be unwise to conclude from this that the earlier debate had been irrelevant. On the contrary, we have arrived at the positions that we now (however transitionally) occupy in part by virtue of having sought to work that discussion through, and of having thus been brought to believe that the question that it may always be relevant to push at appropriate moments is not so much 'Is this

an empirical or a conceptual matter?' as 'Should this matter be treated in this context as depending on primarily empirical or conceptual considerations?'.

We are, hopefully, all of us much more aware than we may have been to start off with of the ways in which the uncertain delicacy of this interplay between the 'conceptual' and the 'empirical' is bound to render any working distinction between them always provisional and, in the last resort, not fully determinable. It follows from this that the ways in which we order our concepts is bound to impinge on what we may take to be the outcome of observation and experiment when we come to test our theories against the 'reality' that we are trying to identify, to understand and to explain. It may also follow that philosophers, as they work primarily on their analyses of concepts, and scientists, as they work primarily on their investigations of 'reality', have more regular and thoroughgoing need of each other's participation than present institutional habits and arrangements can easily provide for.

A second and not altogether unrelated issue that featured frequently in our earlier discussions turned around questions concerning the classification of different forms of behaviour. Is it in principle possible to draw lines between goal-directive and non-goal-directive, or more specifically intentional and non-intentional, behaviour so workably clear-cut as to enable one to say, on the basis of the most detailed observation, of any given piece of behaviour that it fell fairly and squarely into either the one class or the other? If there is in principle any way of so classifying behaviour as to achieve this result, we certainly did not find it. Perhaps it is too early for such a venture to succeed; maybe we need much more 'hard' scientific information before the basis of any such classification could be constructed. But it may also be that intentionality does not reside in particular strictly observable forms of behaviour at all, in particular kinds of feed-back loops or in certain characteristic sets of equations. At all events, even if problems of classification have not altogether disappeared, we now find ourselves much more inclined to ask not so much '*What* precise forms of behaviour, if any, are intentional?' as '*Why* are we led to make use of such intentional concepts as we do, and do we really need them for the satisfactory characterisation and explanation of certain human and perhaps other animal forms of behaviour?'

There is a third cluster of problems about which it is harder for us to be sure whether or to what extent they may have survived

as a source of potential confusion for ourselves and our readers. Every specialist professional group — psychologists, physiologists, philosophers or whatever — are bound over the course of time to develop their own special vocabularies, their own bodies of authoritative texts, to which compressed reference can easily or even 'must' be made, their own technical procedures, their own peculiar use of otherwise quite common words. (Not to mention the fact that only too often members of one and the same family of specialists may use the same words in significantly different ways, the differences being rooted, as often as not, not merely in careless discrepancies of surface usage but in deep underlying differences of theoretical analysis and understanding.)

Inevitably, quite a large part of the seminar discussions between partners coming from such different disciplines was devoted simply to explaining ourselves to each other. In writing our first chapters for this book we have tried to disentangle ourselves from avoidable misunderstandings between ourselves, while yet making use of that earlier experience to enable newly participant readers to avoid falling into similar misunderstandings. How far we may have been successful in achieving this aim is for us hard to tell. On the one hand, we may have gone so far in taking for granted the elimination of sources of earlier misunderstanding as to leave them in effect as unmarked booby traps for the unwarned reader, relatively unfamiliar as he or she is almost bound to be with at least one or other set of our originally disparate assumptions. On the other hand, we may, almost certainly and despite all our previous efforts, ourselves have persisted in certain mutual misunderstandings. Some of these, indeed, actually emerge on the explicit surface of our second chapters; others, no doubt, remain for the reader to detect, if he or she can.

Part of these difficulties lies in a phenomenon of which we have become increasingly aware as we have gone along. We have already noted the impossibility in principle of establishing any hard and fast overall boundary between the empirical and the conceptual. But this difficulty is not wholly independent of one that we have noticed in trying to keep track of our own mutual disagreements and in determining the extent to which they may be 'merely' terminological or, on the other hand, substantial. In so doing we have become aware of a tendency, perhaps more natural to scientists than to philosophers, to soften the threatening outlines of looming substantial disagreement by way of an implicit mutual agreement to treat it as one of merely discrepant terminology. We

would urge our readers to be similarly aware of this temptation and of this problem.

If this is a temptation that comes naturally to scientists, it may be because of the common and understandable assumption that the natural sciences are to be thought of as together contributing to one internally coherent account of the universe as a whole, and that if, therefore, two scientists actually disagree on a matter of substance, one of them must be wrong. Not all philosophers would feel themselves so sustained or constrained by any comparable view of the 'objectivity' of their branch of learning. Be this as it may, readers of this book should be warned that they must not take it too easily for granted that all of the contributors have always succeeded in using the same terms in the same consistent way as each other, even when they appear to be most directly in mutual agreement or at mutual loggerheads. Indeed, the interplay between fact and terminology is one of the most fascinating and tricky aspects of this whole area of debate.

We have already noted that in the course of the discussions to be found gathered together in this volume we have had, scientists and philosophers alike, both to spell out certain things that we should not normally feel the need to spell out at such a level of debate with 'mere' fellow professionals, and yet also at times to limit ourselves to highly abbreviated and virtually unargued statements of our own particular views on what are in fact complex and controversial matters. We have also had to leave more or less unmentioned whole bodies of debate on topics closely interconnected with those at the heart of our present concerns, but which for one reason or another have not come to the forefront in our discussions. So far as the primarily philosophical literature is concerned, one may think, for example, of the debate surrounding the mind–brain identity thesis, of the discussions concerning the analysis of such notions as 'function' and 'role', of the arguments that have been presented both for and against the hypothesis of a so-called 'language of thought', of the various theories that have been put forward and attacked as to the best way of understanding the relations between the concepts of 'reason' and 'cause' (and their analogues in other languages). There are also, of course, topics in what more generally comes under the title of moral philosophy, such as that of 'the freedom of the will', and related problems concerning the nature of responsibility, which likewise touch very closely on the issues under present discussion and around which a vast literature has built up.

It is not that such topics are not touched on, if not explicitly then at any rate implicitly, in the discussions that follow; but we should say clearly that we make no pretence of trying to engage with the extensive established literature that has been devoted to them. (Reference to that literature can, however, be tracked down by judicious use of the suggestions for further reading and by the bibliography provided on pp. 295–301, either directly or, more often, by onward reference from the works that are listed in one or the other.)

Our main aim, then, has been to forward discussion of the matters of our common concern, each with colleagues of other disciplines as well as with those of his or her own. Consequently, we should on the whole expect those of our readers with prior experience of the philosophical literature to find greatest immediate interest and stimulus in the contributions by the scientists amongst us and *vice versa*; though, from our own point of view, we take the main interest of these debates to lie in how each reacts to the other, and in the directions in which the subject is taken when philosophers and scientists of different convictions and experience discuss it seriously together. What should not be expected, however, are scholarly articles of the sort for which one might very properly look in the distinctive specialist journals of psychology, philosophy, physiology or animal behaviour. We have not here sought to advance the 'strictly professional' aspects of our subjects in that way.

But let not this disclaimer be misunderstood. We *do* believe that in the longer run each of our own professionalisms stands to benefit in essential ways from serious exchange with each others'. That is to say, we do *not* believe in the long term viability of strictly compartmentalised approaches to the understanding of human nature and the human situation.

There is another general point to which it may be helpful to draw preliminary attention. This introduction opened with a declaration of intention, a declaration that cannot, one might think, be understood other than on the supposition that those who issue it take their intention to have had at least some effective guiding influence on their production of the texts that follow. Physiological psychologists and animal behaviourists would, in the present state of the art, regard it as being an already very considerable achievement to be able to provide sufficient explanations of even relatively simple instances of apparently goal-directive behaviour of relatively simple organisms such as bees or spiders or

birds. (Such organisms are already complicated enough.) Here, there are three tendencies: either actually to define those purposive concepts which one might want to apply to such behaviour in terms of those special patterns of causal interaction that seem to explain its occurrence; or to maintain that the whole concept of an intention is confused and unnecessary; or to argue that whatever intentions we may ascribe to ourselves or to others in fact play no part whatsoever in the production of the behaviour that we observe. (This is the tendency so forcefully expressed by David McFarland in his contributions to this volume.) The hope is then expressed — indeed, it may be held to as a matter of scientific faith — that in the long run all observable behaviour, including linguistic behaviour and even such linguistic behaviour as may be made manifest in the self-ascription of intentions, may be shown to be explicable in the same general ways.

Those who feel most uneasy and uncertain as to what sense may be made of such claims tend to start from the opposite end of the spectrum, and hence to argue in characteristically different ways. That is, they tend to *start* from a consideration of linguistic behaviour, from an attempt to show in what ways concepts of intentionality may be indispensable to any adequate characterisation of it (and hence to any adequate explanation of its occurrence), and then to ask whether it may perhaps turn out to be more appropriate to characterise certain forms of *non*-linguistic behaviour in nevertheless analogously intentional ways. (Alan Montefiore provides a clear example in this volume of this form of argument.)

Thus, there are at least two clearly distinguishable ways of approaching the problems. One — which may be called, more provocatively than accurately no doubt, the 'normal' scientist's way — is to start by working with familiar tools on such relatively simple areas as are most amenable to immediate study in the hope that the same methods and patterns of explanation may subsequently be extendable to all other cases, including even the most complex. The other — which, in a spirit of similar provocation and disdain of strict accuracy, may be called the more typically philosophical — is to start by reflecting on the worst, i.e. the most complex, case in an effort to establish, inevitably by conceptual rather than by experimental means, what are the minimum conceptual conditions that any adequate account must satisfy; and then to speculate on whether there might turn out to be good reason to extend the use of whatever concepts might be found to be indispensable in the worst case back down the spectrum, so

to speak, to help in the better understanding of cases in which it might have seemed more possible to get by without them.

What is striking in the discussions recorded and pursued in this book is the way in which both 'sides' to this aspect of the debates have tended to move not perhaps towards agreement, but at any rate towards taking better account of each other's primary concerns; and how, in partial consequence no doubt, questions of linguistic behaviour have taken on an increasingly central role in the discussions between them.

A word should also be said about the apparent absences from our discussions. Why are there no contributions from a physicist, an expert in artificial intelligence, a psychoanalyst, for example? Part of the answer is that physicists *have* taken part in our discussions at more than one point; that while the artificial intelligence specialists whom we had hoped would be able to join us found themselves in the end unable (for practical reasons) to do so, more than one among the contributors has had more than a passing experience of the theoretical as well as the practical world of computers; that at one stage, indeed, we even had a practising psychoanalyst taking part in our meetings (and at another stage an ex-psychoanalyst now turned philosopher). Another and equally important part of the answer lies in the reminder that these discussions are simply what they present themselves as being: that is, the summing up and continuation of an on–going debate between a particular group of people, who feel the need, at the stage at which they now find themselves, to open up their debate to a wider range of participants than those to whom they happen to have easy personal access. Certainly we hope that there may be physicists, artificial intelligence people, psychoanalysts, etc. among this wider range.

If we once again mention what may fairly be called the immensity and immense variety of the specialist background to our own discussions, it is not exactly to apologise for not having tried to bring it into more explicit play. Indeed, had each of us tried first, or at the same time, to address himself or herself to his or her fellow physiologists, psychologists, philosophers or whatever, it is more than doubtful whether our own cross-disciplinary discussions could ever have got going at all. Still, we cannot but remain conscious of how much better equipped we should all be to engage in discussions of this kind of complexity and importance were each of us able to speak (and to think) out of a full and confident acquaintance with the literature and techniques of each other's special disciplines. For many practical reasons this may be a largely impossible ideal, no doubt; it is, for all that, one which it is salutary, perhaps even necessary, to keep before us.

CHAPTER 2

PHILOSOPHICAL BACKGROUND

Alan Montefiore

Of the three philosophical contributors to this volume, two have written contributions which present themselves as being essentially self-explanatory, that is to say as standing in no need of any especial introduction. The ways in which they do this are to some extent different. In presenting her own explanation and discussion of the nature of the controversies in which we are all engaged, Kathy Wilkes manages at the same time to explain the terms of her own presentation; Daniel Dennett, on the other hand, finds ways of presenting and discussing the problems, as he sees them, which hardly appear to rely on or to presume any special previous philosophical experience whatsoever. I myself am thus largely alone in finding myself here relying on, even abbreviating whole stretches of my argument into, what are in effect exceedingly condensed allusions to the history of philosophy. Those who are already familiar with that history are unlikely to have any difficulty in picking them up, even if they may not necessarily agree with my reading of this or that author, this or that thesis or argument. For readers who are not familiar with that history, however, who may even have no prior knowledge of it at all, I should try to provide some clue as to how these allusions are to be taken. It goes without saying that no clue to the understanding of what are by any standard complex and controversial matters can be itself uncontroversial. Often, no doubt, the degree of controversiality of a matter may be taken as one measure of its importance. Moreover, the simpler and sparser the account, the more controversial it is likely to be. No matter; it is better to say something, however brief and simplified it may be, than to provide no clue at all. For

those who are interested to discover more, there is plenty of ancillary literature to help them on their way; and even the original sources themselves often make a much better read than many may have supposed. For immediate purposes, however, only the barest of pointers must suffice.

There is one other preliminary point, however, which may also be worth making. The great majority of contemporary analytic philosophers have probably been brought up to take what is known as a 'problems approach' to philosophy. That is to say that they would find it, as it were naturally, conducive to clarity to distinguish between the articulation, analysis and discussion of philosophical problems on the one hand and, on the other hand, the history of their discussion by previous philosophers. Reference to what the greatest of previous philosophers have had to say may, no doubt, be helpful to us in our own discussions, but it is in principle no more indispensable to their successful pursuance than is a knowledge of the history of mathematics or physics to the successful teaching, learning or pursuit of those subjects. There is another philosophical tradition, however, from whose perspective the idea that philosophical problems exist 'timelessly' or outside the context of their own history, so to speak, may sometimes have some provisional pragmatic utility, but in the end involves always an illusion. For a variety of reasons, into which it would be inappropriate to enter here, I find myself in increasing sympathy with philosophers writing from this perspective; and so, speaking simply for myself, I should be inclined to see my attempt to provide some brief background unpacking of my own historical allusions as being of some potential relevance also to my colleagues' much less historically self-conscious contributions. But *that*, as I say, is an argument into which there is no call to enter here.

We may start, then, with the man whose work is often regarded as marking the birth of modern philosophy, René Descartes. In the first paragraph of page 60 I refer to the experience of 'a present which in some not fully Cartesian sense it must nevertheless deem to be undeniably its own'. What is this sense?

Descartes is famous for having sought to base the whole ordered structure of our knowledge on foundations of such certainty that no-one who contemplated them could fail to recognise them as being secure beyond all possible doubt. Such a foundation he found in the utter certainty which anyone whose attention is turned to the matter must acknowledge in his or her own existence as a consciously thinking being at each present moment at which

he turns his thought thus back upon himself. 'I think, therefore I am.' Everything else, however subjectively convinced I may feel of its truth or reality, is in principle open to some kind of doubt or another. Let me try to doubt my own present existence, however, and the very gesture or act of my own self-conscious doubt, as I turn to reflect upon the problem, provides me with the irrefutable proof of its own baselessness. Such certainty cannot, on the other hand, attach to any thought that I may entertain of either my past or my future. Probable as short-term memory or prediction may be, they are, notoriously, not infallible. Similarly, I can never be one hundred per cent sure of the real existence, independently of my own apparent consciousness of them, of anything or anybody else of which or whom I seem to be aware. Of my own immediately present existence as a (self-)consciously thinking being, however, I can be utterly and indubitably certain. The 'fully Cartesian sense' in which one's own present existence is undeniable is that in which it may thus seem to be directly and unproblematically present to one's own immediate consciousness.

This Cartesian thesis is undoubtedly one of great power. Embedded within it, moreover, are a number of further equally powerful implications. One of these concerns the fundamental nature of our own temporal experience of ourselves — as, indeed, of everything else of which we may seem to become aware as falling within our field of consciousness: namely, that this experience, whatever it may happen to *feel* like, has to be thought of as essentially discrete.

The line of reasoning which leads to this conclusion can be indicated in very simple terms. As we have just noted, memory and prediction are both in principle fallible. It follows that I can never be one hundred per cent sure *either* that I have had a past — maybe God has just created me, or maybe I have simply just come into existence, at this immediately present moment, along with whatever may be my present thoughts, apparent memories and all — *or* that I shall experience a future. My present experience of myself presents me with a striking contrast to these uncertainties. It alone cannot be doubted, inasmuch as it stands by itself. It is, that is to say, whatever it is, independently of whatever the past or the future may have been or may be, independently indeed of whether there has been or will be any past or future at all. And every present moment of experience is necessarily in the same case. As Descartes himself said in his third Meditation, 'every moment of our lives is independent of every other moment'.

The British Empiricists — according to the well-established roll-call, Locke, Berkeley and Hume — in effect accepted, if with varying but increasing degrees of consistency, the two following Cartesian assumptions: (i) that all 'direct' experience is limited to the contents of our own conscious awareness, and, (ii) that each such moment of experience is strictly independent of whatever might have occurred or might occur at any other moment = that is to say that each moment of experience is complete in itself and is recognisable as such. To these they added the characteristically empiricist assumption that all the contents of our mind, all our perceptions or ideas or impressions as they were variously called, have to be understood as either constituting some 'original' aspect of experience, as thus construed, or as derived from it.

It was Hume, it is broadly fair to say, who, of the classical British Empiricists, pushed the consequences of these assumptions to their furthest limits, seeking to display the origins of even our most complex organising ideas in the 'impressions' of immediate experience and in what he took to be the natural tendency of the human mind to form habitual associations. 'Ideas' Hume took to consist of either simple copies of the 'original impressions' or more or less complex juxtapositions of such copies. In the case of 'causation', which he took to be a quite notably complex idea, and which is of course of particular relevance to our own present concerns, he sought to show its origin to lie essentially in the repeated experience of sequences of events (or, as he himself more often said, of objects) following each other with such thoroughgoing regularity as to set up settled habits of expectation that such regularities of succession would continue in the future — that is, that whenever the first member of such a sequence occurred, the second would duly follow.

It was, Hume thought, absurd to suppose that one could ever experience, as a matter of observable fact as it were, anything that might conceivably serve to link the first event, A, of such a sequence (as cause) to the following event, B, (as effect) by a tie of any greater necessity than the 'mere' *de facto* regularity of the sequence and our consequent psychologically accompanying feelings of settled expectation. For suppose that one did come to register a further impression of some sort as presenting itself with equal regularity between the occurrences of some such sequentially linked As and Bs. Given that every moment and item of experience has still to be taken as being fully independent of every other, that our perceptions are all, as Hume put it, 'separate and distinct

existences', what else could the regular experience of any such intervening impression show other than that this particular regularly occurring sequence of essentially discrete but contingently associated events contained more members than supposed?

If the nature of causal connection is thus to be understood as consisting at bottom in nothing more than contingent regularity, in what Hume called 'constant conjunction', and causal explanation of particular events as consisting in their being placed in the context of whatever appropriate regularities, the Humean view that there is simply no incompatibility between the discourse of intentional action, of free will and of moral responsibility and that of thoroughgoing causal explicability or causal determinism, follows as a matter of course. For, as he points out, the freedom of intentional choice stands opposed to compulsion or constraint, while the opposite of consistent regularity in the onward course of events lies in mere randomness or chance. Mere regularity in the repeated occurrence of sequences of in themselves wholly independent events can obviously neither constrain nor compel any one event to occur rather than any other. Indeed, such regularity of *de facto* connection between the occurrence of desire or decision and that of the action desired or decided upon, far from being incompatible with deliberate or responsible agency, is, Hume points out, a necessary condition of it. For what sort of responsibility or even sanity could one attribute to an agent whose actions stood in no relation of regular connection to his own acknowledged desires or intended decisions?

Modern Humean-type analyses of the nature of causal explanation and associated Humean-type doctrines of the compatibility between causation and free intentional action have, of course, become extremely sophisticated. What I have provided here is but a thumb-nail sketch of their ancestry and their rationale. However, it may, I hope, help to clarify the sense of my allusions, on page 61 and again on page 78, to 'Humean compatibilisms'.

It is, inevitably, far harder to provide any acceptable clue — acceptably clear and yet acceptably brief — to the sense of my allusions to Kant (starting with that on page 61); for Kant's writings are, notoriously, among the most difficult of any philosopher of the Western tradition. In sketchy outline, however:

Kant saw that there had to be something wrong with the programme of showing how all our ideas are to be understood in terms of their derivation from an experience consisting of nothing but essentially discrete perceptions. Indeed, Hume himself had

virtually come to realise this, as he makes clear in his Appendix to his great *Treatise on Human Nature*, where he looks back on the analysis that he had attempted in the main body of his text of the idea of personal or self-identity, and recognises that it will not do. Hume, however, caught as he was within a framework of what seemed to him to be clearly unrenounceable assumptions, could see no way out of the impasse in which he nevertheless acknowledged himself to be. For if the series of perceptions, the series of which the whole of our conscious experience is made up, is really one of radical discontinuity, from what element within it can our ideas of continuity, and most notably that of a self or subject of experience, continuous or identical with itself over the time of its own experience, ever arise?

The reply that Hume himself had attempted in the main body of his *Treatise* was that such ideas must have arisen as a result of an illusion, rather in the way in which the cinema-goer experiences an illusion of continuous movement as a result of having a series of discrete stills projected onto the screen before him with what Hume would have called an 'inconceivable rapidity'. The problem with this solution, however, is that it, the problem, immediately recurs. The very account of how such an illusion arises makes implicit but indispensable reference to some idea of an observing subject who must, of course, be one and the same (that is, continuous with itself) throughout the time of its observation of the series of discrete but rapidly successive stills. If there was no such subject, continuously identical with itself from one moment of its experience to another, there would be nobody or nothing in whom the series might produce the illusion of being itself a continuity. Only an already self-identical subject, one might say, could experience the alleged illusion of its own continuous self-identity.

Hume, as I have said, saw, or largely saw, the problem, but could envisage no solution to it. (He did express the hope that he might be able to work it out at some later moment; or thought, alternatively, that it might be something that he would have to leave to his successors.) Kant, on the other hand, may be understood as having made the radical move of taking the reference to a unitary and unifying subject to be a necessary *presupposition* of any meaningful experience whatsoever. Such a subject must be taken as unitary in the sense of being one and the same throughout the whole extent of its own experience, and unifying in that it is to be presupposed as capable of holding the different elements of that experience together in relation to each other; in so