

MECHANISM AND RATIONALITY: THE CASE FOR EXPLANATORY  
INCOMPATIBILISM.

---

Francis X. Williamson

University of Cape Town

A dissertation submitted to the Faculty of Social Science and  
Humanities, University of Cape Town, in partial fulfilment of  
the requirements for the Master of Arts degree in Philosophy.

January 1988

The University of Cape Town has been given  
the right to reproduce this thesis in whole  
or in part. Copyright is held by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

ABSTRACT

Title : Mechanism and Rationality: The Case for  
Explanatory Incompatibilism.

Candidate: Francis Xavier Williamson

This thesis is an attempt to defend explanatory incompatibilism, the view that mechanistic and intentional explanations of behaviour are incompatible, against various sorts of objections which come in the form of rival compatibilist theories.

In the first chapter the author outlines the prima facie case for explanatory incompatibilism. This prima facie case is then bolstered by a discussion of explanation in general, conditions of compatibility for different explanations of the same phenomenon, and then a more rigorous account of mechanistic and intentional explanations which allows for a formal presentation of an argument for their incompatibility.

Chapters Two, Three and Four discuss some of the compatibilist theories which have been advanced.

Chapter Two involves a discussion of the "Double-Language" version of compatibilism as advocated by Ryle and Melden.

This version is rejected for two main reasons:

- (1) it fails to keep the two sorts of explanation sufficiently apart so as to render them compatible, and
- (2) it fails to show that intentional explanations are not a

species of causal explanation.

Chapter Three attempts to deal with the "Instrumentalist" version of compatibilism as advanced by Daniel Dennett. This is rejected because it fails to provide a rich enough account of rational action and it also leads to epiphenomenalism.

In Chapter Four the author discusses the "Physicalist" approach to the question of compatibility as advocated by Alvin Goldman and Donald Davidson. But this version of compatibilism is found to be wanting because it also leads to the epiphenomenalism of the mental.

Chapter Five, the conclusion, summarises the basic argument and attempts to develop the author's own account of what the necessary and sufficient conditions for intentional action are. This is found to involve three main elements: physical indeterminism, intentional intelligibility, and then something like the concept of agent-causation. In the course of this account there is a brief discussion of the problem of other minds and an argument against the desire-belief model of action and its explanation based on its inability to cope with the problem of deviant causal chains. It is concluded that mechanistic and intentional explanations are indeed incompatible and something is said about the broad metaphysical view which is required to accommodate this fact.

DECLARATION

I hereby declare that this is my own unaided work.

It is being submitted to the University of Cape Town for the degree of Master of Arts and has never been submitted to any other university for any degree or examination.

Francis Xavier Williamson

Cape Town

January 1988

## TABLE OF CONTENTS

INTRODUCTION	I
CHAPTER ONE : EXPLANATORY INCOMPATIBILISM	
1.1 Mechanism and Agency	5
1.2 Explanation	13
1.3 Mechanistic and Intentional Explanations	25
CHAPTER TWO : THE DOUBLE-LANGUAGE STRATEGY	39
CHAPTER THREE: INSTRUMENTALISM	
3.1 Intentional Explanations as an Explanatory Device	62
3.2 Designed and Enacted Rationality	71
3.3 Two Sorts of Intentional Explanation	80
CHAPTER FOUR: PHYSICALISM	87
CHAPTER FIVE: CONCLUSION	
5.1 Summary of the Argument	114
5.2 Actions, Agents and Other Minds	117
BIBLIOGRAPHY:	136

## ACKNOWLEDGEMENTS

I wish to acknowledge and thank the Human Sciences Research Council as well as the UCT Bursaries Office for providing the financial assistance required for me to complete this thesis.

I would also like to thank my supervisor, Dr Augustine Shutte, for his gentle, often silent but ever-present support. He not only led me into this field but in his own quiet way encouraged me not to abandon the broad metaphysical view required to accomodate the main conclusions of the thesis. At times it seemed impossible to me, but never to him. Thanks.

And then there is Fiona, Graham, Edwina and Rita, all of whom helped me in so many unseen ways. Thanks.

## INTRODUCTION

The central core of any libertarian conception of human action is and always has been the view that physical indeterminism is somehow required if we are to be free agents. But it has also always been obvious that mere indeterminism is not enough - if it were then our actions would be as free as the random meanderings of some quantum particle (assuming of course, as I will throughout this discussion, that quantum phenomena are genuinely rather than a merely epistemically indeterministic), and this is not a freedom worth wanting, and certainly not a freedom upon which to base our attributions of moral blame and praise.

One recalls A.J. Ayer's classic statement of the dilemma faced by libertarians:

Either it is an accident that I choose to act as I do or it is not. If it is an accident, then it is merely a matter of chance that I did not choose otherwise; and if it is merely a matter of chance that I did not choose otherwise, it is surely irrational to hold me morally responsible for choosing as I did. But if it is not an accident that I chose to do one thing rather than another, then presumably there is some causal explanation of my choice: and in that case we are led back to determinism (1982:18).

In the light of this dilemma, many believers in free action have opted for some or other version of compatibilism, the view, namely, that our being free and responsible agents is compatible with each and every one of our actions being causally determined by antecedent states of the world, whether these be mental events in our minds or physical events in our brains.



But some incompatibilists, those who believe in free action and also believe that this is incompatible with physical determinism, have been unmoved by this. Instead they have attempted to pass through the horns of Ayer's dilemma by developing an account of free action which espouses physical indeterminism but nevertheless shows that our actions are not simply random, but that they in fact are intelligible and explicable in the light of teleological or purposive ways of understanding and talking about human behaviour. But it is not that free action is thought of as being physically undetermined and also teleologically intelligible, but rather that it in some sense has to be physically undetermined in order to be so intelligible. The view is that physically deterministic ways of understanding human behaviour are incompatible with that behaviour also being understood as the purposive and rational activity of agents who can be held morally accountable for their deeds.

The core of this entire dissertation is to defend the view that mechanistic explanations of behaviour, at least insofar as they are complete explanations, are incompatible with that behaviour being thought of as rational action.

Following Gary Watson (1982:12) I refer to this view as explanatory incompatibilism as opposed to the older modal incompatibilism which seeks to show that freedom is opposed to causal necessitation.

In the first chapter I present what I take to be the prima facie grounds for asserting explanatory incompatibilism, and I then go on to give a brief account of explanation in general and the various ways in which different explanations could be compatible. I also offer an account of what

mechanistic and intentional (teleological or purposive) explanations are and I present a formal argument for explanatory incompatibilism.

The following three chapters - the core of the discussion - are all attempts to defend explanatory incompatibilism against certain sorts of objections. The conclusion is reached that none of these objections stand up to critical scrutiny, and that they all, in one form or another, can only show that mechanism - complete mechanistic explanations of all our behaviour - is compatible with 'free agency' by denying that our mental lives make a causal difference to the way our bodies behave.

In the conclusion I present the germ of a view of how physical indeterminism can accommodate the fact, if such it be, that our actions, far from being chance-like events, can be the responsible actions of free and rational agents.

I do not address the complex issue of moral responsibility, action and blame in any direct way, for the discussion is rather concerned with the formal compatibility of mechanism and agency, but the relevance of the view I develop to these issues will be readily apparent.

Needless to say, there are many issues which bear on this complex topic which have been ignored or left out entirely. But if the core of the thesis is correct, as I am convinced it is, then I see no way of denting these libertarian and explanatory incompatibilist claims. Mechanism is a threat to our view of ourselves as free and responsible agents, and if that view of ourselves is to survive, then it can only do so

on the condition that mechanism is false. I hope to show in this dissertation that there are good grounds, good philosophical grounds, for believing that it in fact is false.

## CHAPTER ONE

### EXPLANATORY INCOMPATIBILISM

#### 1.1 Mechanism and Agency.

Common sense, it seems, has always found something in the experience of our own rational agency which is incompatible with the view that we are mere mechanisms, in however extended and complicated a sense this may be.

Underlying this vague but pervasive conviction is the view, or cluster of views, that the verification of a comprehensive mechanistic theory of all behaviour would in some way be in conflict with our view of ourselves as responsible agents who are the bearers of values, choices and goals. Would not all talk of agents, actions and responsibility be relegated to the scrapheap of discredited and outmoded ideas in much the same way as advances in science have cast to the wayside concepts such as phlogiston, vital spirit and the ether? As Stephen Toulmin has expressed the worry:

For, then, every step forward in the neurological analysis of brain-mechanisms will shrink the area of operation within which mind and thought, mental and rational categories, have any application. Like the God of the natural theologians, the Human Reason will find itself on a diminishing sandbank, with the tide of Science rising all around it (1970:1-2. His emphasis).

One obvious way in which the thesis of mechanism threatens responsible agency - and this emerges clearly from the vast literature on free will - is in the formidable deterministic implications which such a comprehensive mechanistic theory

1  
would seem to hold. And even if such a mechanistic theory  
should turn out to be fundamentally indeterministic or  
probabilistic - as it has every chance of doing - there<sup>2</sup>  
remains the problem of justifying our attributions of moral  
responsibility and its allied institutions when it is a  
matter of chance that a person chose or acted one way rather  
than another.<sup>3</sup> But a further and perhaps more fundamental way  
in which mechanism threatens agency is that it seems to  
undermine the applicability of the entire framework in which  
talk of action, agency and responsibility makes sense. For  
it seem that whenever a mechanistic account of some  
particular piece of human or animal behaviour is available,  
then any intentional or purposive explanation is rendered  
otiose or ruled out altogether.

Suppose we seek an explanation for a piece of human behaviour  
which we have characterised as the raising of an arm.<sup>4</sup> We

1. I do not wish to boldly assert that determinism is incompatible with free agency, but only call attention to the fact that it is pre-philosophically perceived to be such. Besides this, even though the whole 'compatibilist' debate is far from conclusive, there is at least one sense of free agency, viz. freedom of indifference, which is incompatible with determinism.
2. No pun intended. But see for instance Suppes 1974 as well as the numerous articles which claim that quantum indeterminacy is here to stay and that it is likely to affect our basic account of brain-function. See Popper and Eccles 1983; Eccles 1984; Swinburne 1986 and 1986A.
3. This problem has been almost universally acknowledged if one is to judge by the frequency of its mention in the literature. Van Inwagen (1983) calls this the Mind Argument for compatibilism because of the frequency of its appearance in that journal. See Ayer 1982:18; D.J. O'Connor 1971:57-58; Chisholm 1982:27-28; Watson 1982:1-14; Flew:117-120.
4. This is an adapted version of essentially the same example used by Dennett 1973:159.

would normally proceed by citing the motives, reasons or purposes the agent had in behaving as he did, assuming, of course, that it was not raised by someone else or by the agent's other arm.<sup>5</sup> We might say that he was signalling the waiter, or waving to a friend, or brushing away a fly, in which case we would have offered an intentional explanation of the behaviour, regarding it as the purposive action of an agent. But suppose a neurophysiologist informs us that in fact the extensor muscles in the man's arm were contracted by nerve impulses triggered by a freak epileptic discharge in the man's brain. We would, no doubt, abandon our original intentional explanation of the behaviour and come to regard it as a mere event or movement, as something which happened to the man rather than as something which he did.

The mechanistic explanation of the behaviour seems to disqualify or displace the intentional or purposive account; indeed, it seems to disqualify the behaviour from being thought of as an action at all, and this is because the behaviour has apparently been shown to bear no relation - neither logical nor causal - to any purposes or intentions of any agent whatsoever, but is simply a motion of a portion of his body.

It might naturally be objected that a simple example like this can be notoriously misleading, for it seems quite obvious that movements caused by epileptic discharges and the like are not plausible candidates for characterisation as purposive action. We should, in all fairness, consider the

5. Assuming, that is, that it is a 'basic action' and not something done by doing something else. See Danto 1965.

case where the mechanistic explanation is far more detailed and complex, stretching as it were into the higher and remoter - perhaps more 'rational' - regions of the brain. We should consider the case where the scientists hand us twenty volumes of fine-print detailing neuronal pathways and cortical structures - "Here, this is how and why the arm went up"<sup>6</sup>. What then? Would this incline us towards the conclusion that the behaviour after all can be the intentional action of an agent? It is difficult to see what difference this would make. At least it is not obvious that adducing more detailed and complex physical causes changes the moral of the example. In either case the behaviour has apparently been shown to be the inexorable (or random) outcome of a series of mechanistic causes which leaves no room for the causal efficacy of the reasons or intentions of any agent and it seems intuitively certain that these must be causally efficacious in bringing the behaviour about if it is to be considered as genuine action.<sup>7</sup>

It is precisely in this sense that we do not regard missiles, clocks and computers as agents capable of reason and purpose who can be held morally accountable for their actions, and the reason is not that they are somehow incapable of acting otherwise than they in fact do - they could just as well be thoroughly indeterministic systems - nor is it simply that they lack that mysterious property called consciousness. The reason rather is that they simply are not capable of action at all. The whole explanatory framework of purpose, reason, action and agency is in some way inapplicable to these

6. This is what Dennett (1973:173) suggests we do.
7. We shall come back to this point repeatedly.

sorts of systems, except perhaps in a reduced and  
8  
metaphorical sense, because their behaviour proceeds from an  
antecedent series of 'blind' mechanistic causes and not from  
any reasons or intentions which are the mark of genuine  
agency and action. It would appear, then, that the  
verification of a mechanistic theory of human behaviour would  
show the same about us, that contrary to our usual and  
cherished beliefs, our behaviour proceeds from a similar  
sequence of mechanistic causes amongst which our intentions  
and goals are sadly not to be counted, having at most an  
impotent and epiphenomenal existence alongside the  
mechanisms which are the 'real' determinants of our  
behaviour. And this idea is not only counter-intuitive but  
deeply disturbing too. As Thomas Nagel has expressed it, "...  
something in the idea of agency is incompatible with actions  
being events, or people being things .... Eventually nothing  
remains which can be ascribed to the responsible self, and we  
are left with nothing but a portion of the larger sequence of  
events, which can be deplored or celebrated, but not blamed  
or praised" (1979:34).

The problem, of course, is as old as philosophy itself. In  
its most fundamental form it consists in finding a way to  
reconcile what seem to be two incompatible intuitions about  
ourselves and the world we live in: on the one hand there is  
the undeniable fact of our own rational agency, and on the  
other hand is the intuition that as flesh and blood creatures  
inhabiting this physical world there can be no upper limit to  
what a mechanistic science of man might reveal about the

8. Charles Taylor 1964:54-71 and also R. Taylor 1966: 238-  
243.



physical mechanisms which underlie all our behaviour.

Plato can be said to have discussed it in the famous passage in the Phaedo where Socrates, in the course of levelling an attack on the materialism of Anaxagoras, talks about the 'real' causes of his sitting in prison awaiting execution:

As I read on I discovered that the fellow made no use of the Mind and assigned to it no causality for the order of the world, but adduced causes like air and ether and water and many other absurdities. It seemed to me that he was just about as inconsistent as if someone were to say "The cause of everything that Socrates does is Mind" and then, in trying to account for my several actions, said first that the reason why I am lying here now is that my body is composed of limbs and sinews, ... and the sinews are capable of contraction and relaxation, ... and since the bones move freely at their joints, the sinews by relaxing and contracting enable me to somehow bend my limbs; and that is the cause of my sitting here in a bent position ... never [troubling] to mention the real reasons; which are that since Athens has thought it better to condemn me, therefore I for my part have thought it better to sit here, and more right to stay and submit to whatever penalty she orders - because, by Dog! I fancy that these sinews and bones would have been in the neighbourhood of Megara or Boeotia long ago ... if I did not think that it was more right and honourable to submit to whatever penalty my country orders rather than take to my heels and run away (98B-99D:156-157).

Plato goes on to admit that the "sinews and bones" have something to do with his sitting in prison, but denies that such an account could ever be complete:

If it were said that without such bones and sinews and all the rest of them I should not be able to do what I think is right, it would be true; but to say that it is because of them that I do what I am doing, and not through choice of what is best ... would be a very lax and inaccurate form of expression (99D:157).

Plato's 'solution' to the problem, as is well known, was a

9. For an interesting discussion of this passage and its place in the history and development of the mind-body problem see Popper and Eccles 1983:169-170.

radical dualism in which the mind (with its reasons and intentions, etc.) belonged to a different immaterial world which somehow inhabited the body, essentially the same 'solution' that was to emerge with Descartes. Now I do not want to embark on a historical survey of the development of the mind-body problem - enough has surely been written about that - but only offer, in this chapter at least, an account of what seem to be the prima facie grounds for asserting that there is an incompatibility between genuine agency and the possibility of unlimited success in our ability to offer exhaustive mechanistic explanations of all behaviour. And the central idea behind the thesis of incompatibilism is the view that whenever a mechanistic explanation of some particular piece of human behaviour is available, then any intentional or purposive explanation of that behaviour is rendered otiose or ruled out altogether.

Considerations such as these have led some recent philosophers to claim that mechanistic explanations of behaviour are incompatible with intentional explanations of that same behaviour, in the sense that the two sorts of explanation cannot both be held to be true accounts of a single instance of behaviour. Consider Alasdair MacIntyre's claim that "human behaviour can only be understood in terms of such distinctive concepts as purpose, intention, consciousness, rationality, morality and language. And these concepts rule out the possibility of causal explanation in the sense in which mechanical explanations are causal explanations" (1960:91). Or consider his earlier statement

10. See Popper and Eccles 1983; Shaffer 1968; McGinn 1982.

that "to show that behaviour is rational is enough to show that it is not causally determined in the sense of being the effect of a set of sufficient conditions ... (1957:35). A more extensive defence of what has been variously referred to as "the conflict thesis" (Flew 1978), "the rivalry thesis" (Charles Taylor 1985A) or, as I shall call it, "explanatory incompatibilism" is to be found in Norman Malcolm's "The Conceivability of Mechanism" where he argues that the verification of a comprehensive neurophysiological theory of behaviour would "refute" our ordinary purposive or intentional explanations.

I take it, then, that there is a prima facie case for asserting explanatory incompatibilism, and my aim in the next two sections of this chapter is to see what lies behind these claims and to see whether a more rigorous account can be offered of not only what these claims come to, but also of what these two warring sorts of explanations basically are.

## 1.2 Explanation

The previous section sketched what seem to be the prima facie grounds for explanatory incompatibilism. What this thesis amounts to is the claim that when and wherever a mechanistic explanation of some piece of behaviour is available, then any intentional explanation of that behaviour is ruled out or rendered entirely otiose. I have until now spoken somewhat loosely of mechanistic and intentional (or purposive) explanations as though these types of explanation were clearly defined and well understood, so it is now my aim to provide a more specific account of what exactly these two forms of explanation are and how they differ. But before doing that, let me take a brief look at explanation in general.

An explanation might very simply be viewed as any answer to a 'why' question that makes the event or phenomenon in question somehow appear more intelligible (Boden 1978:32). This is usually done by placing the event or phenomenon in question in a context within which its occurrence becomes part of that which is better understood and explained. Given the context, the occurrence of the event in question is somehow rendered natural and obvious. Arguably the best, and the most widely accepted, account of explanation is that of Carl Hempel (1965, 1966). The precise details of Hempel's account of 'scientific' explanation are not important for this discussion - we are interested only in a broad characterisation - so any rival account will not affect the

basic argument. The basic idea behind Hempel's account is that explanation ideally takes the form of a deductive argument, so that from a statement expressing some particular state of affairs or initial conditions and a proposition expressing a universal law of nature, the occurrence of the phenomenon to be explained can then be deductively derived. This is the so-called deductive-nomological or covering law model of explanation. For example, let us suppose that we are to account for the fact that this particular piece of copper expands when heated.

We have our statement of the initial conditions, viz. this metal is heated, and we also have a proposition expressing a relevant universal law of nature, viz. all metals expand when heated. Then with the addition of a further premise - copper is a metal - we can deduce that this piece of copper will expand, which is what we set out to explain.

This would be an example of a causal explanation which conforms to the deductive-nomological model, because the heat is cited as the cause of the copper's expanding, and citing the heat as cause is taken as being explanatory because it is backed up or 'guaranteed' by an empirical generalisation, in this case a law of nature, to the effect that heat always causes metals to expand.

The deductive-nomological model of explanation is generally held up as the ideal of scientific explanation, and there is

11. For a discussion of the problems in Hempel's account see Swinburne 1979:27-30.
12. This formulation is intended to be neutral with respect to any rival theories of causation and explanation which have been offered.

at least one very good reason why this should be so. A deductive-nomological explanation of some particular event renders that event as uniquely necessary given the antecedent conditions and the operation of the universal law, and so it provides a compelling explanation of why that event and no other had to occur. It is the determinist's belief, or hope, that all events are capable of being explained by antecedent conditions which together with suitable deterministic laws of nature logically entail their occurrence.

13

Philosophers do, however, recognise a weaker sort of explanation, what Hempel refers to as inductive-statistical or probabilistic explanation (1966:58-69), which, because it does not rely on a universal law but rather on a statistical or probabilistic generalisation, does not logically entail the occurrence of the event to be explained but nevertheless renders it as likely to occur. The likelihood of the event's occurrence will obviously be determined by the strength or force of the probabilistic generalisation (Hempel 1966:58-59). An example of this type of explanation - Hempel's own example - would be our explaining Jim's getting the measles by saying that he caught the disease from his brother who suffered a bad case of infection just a few days earlier. There is no universal law connecting exposure to the measles with contracting the disease, but there nevertheless is a high probability that persons exposed to the measles will in fact contract the disease. Given our knowledge of this high probability, and

13. This is not far off Laplace's classic formulation of the thesis of determinism. See van Inwagen 1983:1-22.

our knowledge that Jim was exposed to the disease, we are able to infer that there is a high probability that Jim caught the disease from his brother, and this constitutes our explanation.

This is of considerable importance for the debate about mechanism and agency, as it is for the entire free-will issue, because if the most fundamental laws of nature should in fact be indeterministic, as current physics suggests they indeed are, then this would imply that we can only have partial explanations of the phenomena covered by these laws. The explanations which rely on these laws would show us why the phenomena in question were likely to occur, but could never give us certainty as to why they in fact occurred - because the explanations would be logically consistent with the idea that the events concerned were either uncaused or caused by something completely different to that which is cited in the explanation. This can be made clear as follows: a full explanation is one which renders the explanandum event as necessary given the antecedent conditions and the operation of just those universal laws of nature - it shows that given these antecedent conditions and the operation of the universal laws, the explanandum event could not fail to occur. But what makes this type of explanation possible is just that we know of the requisite universal laws. Should we be ignorant of them, either because our science is insufficiently advanced or because there just are not any such laws to be known (as indeterminists believe is in fact the case), then the explanandum event is rendered only as probable (depending on the probabilistic weight of the generalisation) but not necessary. It would be possible for

the initial conditions and the laws of nature, such as they are, to be just as they were before and yet the explanandum event can fail to occur.<sup>14</sup>

The importance of this, I want to argue, lies not in the possibility of their being crucial gaps of randomness in the otherwise rigid causal structure of the physical world through which an immaterial mind or self can be sneaked into the picture, as Eccles (1984) appears to think, but in the possibility it leaves open for our ordinary mentalistic explanations of human behaviour being the best or strongest explanations of these phenomena that we could have. The point is that there is a common assumption among most writers on this topic that so-called scientific or mechanistic explanations of phenomena enjoy a hegemony over our ordinary mentalistic explanations.<sup>15</sup> The reason for this is that our mentalistic language will never provide us with tight laws of human behaviour which could serve as the basis for accurate prediction and explanation, whereas a physicalist language presumably could.<sup>16</sup> But should we be able to have only partial explanations even at the mechanistic or physical level, even with a fully developed and mature mechanistic theory of human behaviour, then it remains an open question whether the mechanistic or intentional explanations will be the best or strongest. It might very well be the case that our ordinary folk-psychological explanations of human actions render the phenomena concerned more strongly than any mechanistic explanations do. That is to say that an explanation of a

14. For a more detailed account of partial and full explanations see Swinburne 1979:25-46.

15. See for instance Dennett 1973 and Charles Taylor 1964.

16. Davidson (1980B) argues for this view at length.



piece of human behaviour in terms of the intentions or reasons of the agent whose behaviour it is might be a fuller and better explanation than any would-be account in terms of neuronal firings and nerve impulses. <sup>17</sup> But these ideas will be developed more coherently as the discussion progresses. Right now I want to say something about the relation that may exist between two or more explanations of a particular event, and this is directly relevant to this entire discussion because it is the explanatory incompatibilists' claim that mechanistic and intentional explanations of a particular piece of behaviour cannot co-exist. And here, as before, the aim is to provide only an outline of what seem to be the most salient points pertinent to this particular discussion. A more detailed discussion would take us too far afield. One issue which is not directly addressed is the question of reductionism, what exactly it is and when it can be effected. There are two reasons for this: (1) much of what will be said can be taken as an implicit discussion of the question, and (2), as a result of the arguments of philosophers like Davidson (1980B,E,F) and Putnam (1974) very few philosophers anymore take seriously the idea that our mentalistic vocabulary can be reduced to a physicalistic vocabulary. For a full treatment of the question of reductionism in relation to psychology one can consult any number of excellent books and articles dealing with the subject, especially Boden 1978 and Fodor 1968.

In his Free Will: A Defence against Neurophysiological Determinism, John Thorp develops an idea which he calls the

17. These issues are discussed in considerable detail by Thorp 1980, especially pp.85-93.

hegemony which one description or explanation of an event can have with respect to any other description or explanation. His idea is roughly that the description of an event under which that event is most strongly explained or rendered is hegemonic with respect to any other description. Let me quote Thorp somewhat at length:

Explanations differ in strength. The weakest merely render an event. Most explanations are of this type: they give reasons why an event happened, but those reasons would also have been reasons for some other event, incompatible with the first, to happen ... Explanations of middle strength render an event uniquely: they give sufficient conditions for an event under that description to occur, rather than conditions which are sufficient merely for a disjunction of kinds of events. At their best, teleological explanations are of this sort. The strongest degree of explanation is that which renders the explanandum as necessary. The most typical case here renders the explanandum as naturally or physically necessary ... Finally, the description under which an event is most strongly rendered by explanation is hegemonic over other descriptions under which it is rendered less strongly, or under which it is unexplained (1980:90-91.)

The idea of hegemony can be illustrated by a somewhat graphic example. Suppose we explain John's death by citing its cause, viz. that he was shot at point-blank range. This is a rather weak explanation, because not all shootings at point-blank range result in death. But suppose we improve the explanation by redescribing the cause as the shattering of John's brain by a high-momentum bullet, then it seems that we would have provided an explanation which is hegemonic with respect to the first because the explanandum event - John's death - is rendered more strongly (more probable) at this level of description and explanation than the former. We might even provide a more hegemonic explanation by appealing to information which only neurophysiologists are likely to

have at their disposal - an explanation which links up certain vital functions of an organism with certain cortical processes and structures. The idea, then, of hegemony is neither strange or new. It is merely another way of describing the phenomenon of certain regularities and laws being themselves explained by yet more fundamental regularities and laws, in much the same way as Kepler's laws are in turn supported and explained by Newton's laws, which have a wider and more systematic application (Swinburne 1979:45; Taylor 1970:50-54).

None of this should be taken as suggesting that our usual attempts at explanation should always be to provide the most hegemonic explanation we can. On the contrary, we often have a right to be satisfied with something less, as the example above illustrates. Explanations are attempts at illuminating the unfamiliar, and they are context-dependent in the sense that the most illuminating explanation of a phenomenon will vary according to the interests and aims of those seeking enlightenment. The explanatory power of an explanation will depend upon its alighting on just those features which are pertinent to our particular interests at the time. Putnam has a famous example in which he contrasts two different explanations of the fact that a square peg a fraction less than 1" across passes through a square hole 1" across in a board which also has a round hole 1" in diameter, through which the peg does not pass. Now we might "by a heroic feat of calculation" deduce from a description of the board, holes and peg in terms of elementary particles and velocities that the peg will pass through the one region and not the other, but the fact that we might be able to deduce this does not by

itself guarantee that anything will have been explained. We might, far more easily and simply, just say that the board and peg are approximately rigid and one of the holes is big enough for the peg to pass through and the other not - and in this case our explanation would have brought out the salient features of the situation which the other explanation did not

18

(Putnam 1974:132). It is in any case not clear which of these two explanations is the hegemonic one - one could plausibly argue that it is the macro-explanation which renders the explanandum events more strongly, because it does seem to follow with deductive certainty that the peg will pass through the square hole and not the round one once the relevant sizes and the rigidity of the objects have been taken into account. The general point, however, is that reducibility does not guarantee that phenomena will be better explained at the micro or lower level than the higher "ordinary" level.

Once it has been admitted that there can be two or more descriptions and hence also explanations of a particular event, the question arises as to what the relations between two such explanations must be if both are to be accepted as valid accounts of the phenomena in question. Swinburne (1979:44-47) argues that there can indeed be two distinct explanations of an event when either of two conditions

18. Putnam is making a point about reductionism in this example, something which my use of it does not quite bring out. His point is roughly that reductionism does not entail that phenomena can be better explained in terms of lower-level laws and principles than higher ones - in fact the lower level laws and principles might be explanatorily irrelevant as far as the behaviour of macro-objects are concerned.

are met. The first condition can be met when either of the explanations or both are partial explanations of the event in question. As an example Swinburne offers the following:

Thus a man's death from cancer may be explained by (1) his smoking and a generalisation about the proportion of smokers who die from cancer, and by (2) his parent's having died of cancer and a generalisation about those whose parents die of cancer who themselves die of cancer. Since 1) and 2) only make probable but do not necessitate the man's death from cancer, they are only partial explanations. Clearly they can be combined into a fuller explanation ... (1979:45).

The second condition Swinburne argues for can be met when the two explanations are systematically related in the sense that the one can account for the regularities and laws appealed to in the other, and this can be true even when both are full explanations of the event in question. This would be the case when there is a relation of reducibility between the two explanations.

For example, the present position of Mars is explained by its position in the last six days and the laws of planetary motion formulated more or less correctly by Kepler ... Yet the present position of Mars is also explained by its position and velocity last week and those of all other heavenly bodies, and Newton's laws of motion ... Both are full explanations, yet they are clearly compatible. This is because Newton's laws and the positions and velocities of the planets explain their (approximate) conformity to Kepler's laws. Kepler's laws operate because Newton's laws operate ... (1979:45).

Charles Taylor uses the same example as Swinburne to illustrate the second condition, but talks instead of one explanation being "more basic" than the other.

Explanation in the strong sense therefore builds in tiers in a way which cannot be understood simply as the subsumption of less under more general laws. Rather the correlations on one level are explained by those on a deeper level in a way which shows their relation to other possible outcomes. The first can be

seen as a special case of the second: they are shown as exemplifications of the deeper correlations under some initial conditions. Thus the laws of Kepler are the laws of Newtonian mechanics exhibited in the relevant initial conditions holding for our solar system. We can thus often say of two explanations of a given phenomena, that they are compatible but that one is more basic than the other, for it appeals to deeper level laws which themselves can be used to explain the laws involved in the shallower account (Taylor 1970:52).

The second condition mentioned here might be thought too strong because it insists on some relation of reducibility or derivability between the two explanations, and this is contrary to what proponents of the token identity thesis of mental and physical states insist on, that even though there are no or can be no laws connecting types of mental states with types of physical states, each and every state nevertheless is identical with some other physical state (Davidson 1980B). To accomodate these intuitions one need only insist that the two explanations are systematically related in at least the sense that the entities and states picked out in the one are identical with the entities and states picked out in the other even though no relation of reducibility or derivability between the two respective sets of laws or languages obtains. The important point is that there cannot be two unrelated full explanations of an event, because a full explanation gives necessary and sufficient conditions for the occurrence of an event, and there cannot be two unrelated sets of necessary and sufficient conditions for the occurrence of an event. As Swinburne (1979:46) argues:

For suppose that causes and reasons F1 and also causes

19. I shall come back to this point in the chapter on Physicalism.

and reasons F2 each fully explain E, and neither in any way explains the other. Then F1 is necessary and sufficient for the occurrence of E. Given F1, E cannot but occur. Now suppose F1 not to have occurred ... Then since F1 in no way explains the occurrence of F2, the non-occurrence of F1 would in no way affect whether or not F2 occurs. So F2 will still occur. Since F2 provides a full explanation of E, it is necessary and sufficient for the occurrence of E. So E will occur, even though F1 does not occur. So F1 cannot be necessary and sufficient for the occurrence of E, and so the causes and reasons involved in F1 cannot fully explain E - contrary to our original supposition.

This then completes the discussion of explanation in general.

The conclusions can be summarised as follows:

- (1) there can only be one hegemonic explanaton of a particular phenomenon.
- (2) if there is more than one explanation of some phenomenon, then the hegemonic explanation will be the one that renders the explanandum event most strongly.
- (3) a full explanation of an event will always be hegemonic with respect to any partial explanation of that event.
- (4) there can be two or more disctinct explanations of an event, but where this is the case then (a) at least one or both must be partial explanations, or (b) they must be systematically related in the sense indicated above, viz. they must be related as the more to the less basic (in Taylor's sense) or there must be a relation of referential identity between the entities and states picked out in both explanations.
- (5) there cannot be two (or more) unrelated full explanations of an event.

### 1.3 Mechanistic and Intentional Explanations

Enough has been said, I trust, about explanation in general and the relations that may exist between two or more explanations of the same phenomenon. I now want to redeem my undertaking at the beginning of Section 1.2, which is to provide a more accurate and specific characterisation of the two sorts of explanation which are held by incompatibilists to be in conflict, viz. mechanistic and intentional explanations of human behaviour.

Philosophers have, of course, used different rubrics to label these two basic types or modes of explanation, but there appears to be a rough consensus as to where the distinction falls. Thus what I term mechanistic explanation<sup>20</sup> has been variously referred to as "scientific" (Hempel 1966; Swinburne 1979), or physical, or just simply (and not unambiguously) as causal explanation (Peters 1958; Melden 1961).

These explanations typically account for behaviour by citing the causally sufficient physical conditions for the occurrence of that behaviour, and the terms and concepts which such explanations employ would typically come from the so-called natural or physical sciences. One would therefore expect items such as chemical reactions, neuronal firings and muscular contractions to feature in a mechanistic explanation<sup>21</sup> of some particular piece of behaviour. The reason these are

20. "Mechanistic" is not to be confused with "mechanical", viz. explanation in terms of the laws and concepts of classical mechanics, using only such notions as force, mass, velocity and position. See Boden 1978:32-39.

21. See Dennett 1973:161.



referred to as mechanistic explanations rather than just causal explanations is that the notion of "causal explanations" is ambiguous: one could hold with Davidson (1980A), Goldman (1970) and McGinn (1979) that explanation in terms of purpose or intention is a species of ordinary event-causal explanation, and also insist that explanation in terms of purpose and intention differs in a number of important respects from our more ordinary notion of causal (i.e. mechanistic) explanation. Alternately, one could hold with Richard Taylor (1966) and Roderick Chisholm (1976,1982) that the proper analysis of action requires a notion of causation by an agent which is not reducible to event-causation. Chisholm speaks of immanent (agent) as opposed to transeunt (event) causation (1982:28). The point I wish to make is that if this notion of agent-causation should ultimately prove acceptable, then one cannot simply talk of causal explanation without distinguishing exactly what sense of "cause" is intended. Thus one might plausibly maintain that explanation by purpose and intention is a variety of causal explanation even though it is not event-causation that is meant. For this reason, then, the notion of "causal explanation" is ambiguous and will be replaced, in this discussion, by the more circumspect notion of mechanistic explanation which marks off clearly what sense of "cause" is intended.

A mechanistic explanation, then, is a causal explanation where "causal" is to be construed as referring to efficient causation or causation by virtue of some antecedent physical event. Two important points - partially referred to above - need to be emphasized :

(1) efficient-causation is always a relation between events

(however described) and never between an object and an event or vice versa. All mechanistic explanations are what has been called event-causal in character (Bishop 1983; Davidson 1980A, B, E, F, H). By this is meant that the causal relation which is said to exist between two phenomena when one is cited as the cause of the other, when it is efficient-causation that is intended, is a relation between two events (however described). When we say that the stone broke the window, then this is taken to mean that some event involving the stone caused the breaking - perhaps its momentum or its velocity, but not just the stone all by itself (Swinburne 1979:31). Causality, the slogan runs, is a relation between events (however they are described). And it is perhaps not inaccurate to say that most philosophers today believe that efficient-causation is the only kind of causality that there is - even Aristotle's final causes are said to work, ultimately, by way of  
22  
efficient causes.

(2) a mechanistic explanation is always couched in terms of relations between physical events. An explanation which cites a mental event as a cause would not count as a mechanistic explanation even though it could be regarded as a causal explanation. This would remain so even though the mental event cited is identical with and can be described as  
23  
a physical event.

A mechanistic explanation will then be any explanation which

22. For arguments for a more generous interpretation of "event-causation" and a comment on Davidson's work in this area see Steiner 1986.

23. See Bennett 1976:48-65 and also pp.72-81.

cites an antecedent physical event or state of affairs as the cause of the phenomenon to be explained, and such an explanation may or may not be supported by or grounded in a generalisation which causally links the events thus described in a lawful way. If a mechanistic explanation conforms to the deductive-nomological model - if it renders the explanandum event as uniquely necessary - then it will not only be a full explanation but also a deterministic explanation.

There are, of course, other types of explanations which are not causal explanation in any straightforward understanding of the term. There are, for instance, what John Lucas describes as formal explanations, which are the sort of explanation we would give when trying to explain why "257 is a prime number, or  $\sqrt{2}$  is irrational, or the monadic predicate calculus is decidable" (1970:48). We shall not be concerned with any of these types of explanation, but shall go on straightaway to teleological explanations of which intentional explanations are a type. Teleological explanations account for behaviour in terms of the end for which the behaviour comes about, and they would typically be based on a teleological law to the effect that whenever a system is in a certain state, it will behave in whatever way is required for it to either maintain that state or achieve another (Bennett 1976:39; C. Taylor 1966:3-25).

The peculiar thing about teleological explanations is that they contain an essential reference to a later time in that they explain an event in terms of its either leading to or being required for a future end or state. It used to be argued

24. For a full account of these other types of explanation see Nagel 1971.

that teleological explanations are "unscientific" or in some sense only pseudo-explanations because they reverse the direction of causality by explaining an event in terms of its "effects" rather than an antecedent cause. The work of philosophers like Bennett, Nagel, C. Taylor and Larry

25

Wright on teleological explanation has, in my view decisively, answered these sorts of objections and shown that they are simply based on a misunderstanding. As Bennett argues, it is true that a teleological explanation "reaches forward in time, not for anything so absurd as a later cause, but for a possible later event which is mentioned in the description of the earlier cause" (1976:41).

Intentional explanations are the explanations which feature in our ordinary and everyday account of human (and a lot of animal) behaviour. These are the explanations which cite the purpose, reasons, beliefs, desires, goals and intentions which the agent had in behaving in that way, and they have been variously referred to as rational (Lucas 1970; Pears 1973), purposive (C. Taylor 1964), personal (Swinburne 1979) or just as intentional (Dennett 1973; Bishop 1983) explanations. Davidson refers to this type of explanation as "rationalisation" because the beliefs and desires we attribute to an agent are such as to make the behaviour appear rational or reasonable in the light of these beliefs and desires (1980A:3), and these are the explanations we normally give of actions rather than mere bodily movements. There are two important features of intentional explanations

25. Bennett 1976; Nagel 1971; C.Taylor 1964; Larry Wright 1976.

which emerge from their discussion in the literature:

(1) intentional explanations are teleological in form in that they explain by citing the end for the sake of which the behaviour comes about (C.Taylor 1964:54).

(2) they are not merely teleological but also essentially refer to some or other intentional state of the agent whose behaviour is being explained. All that is meant by this is that the behaviour is explained by reference to some or other psychological or mental object, whether this be of belief or desire, which the agent had in mind (Boden 1978: 47-50). This is important because an explanation can quite easily be teleological without it being intentional. Charles Taylor distinguishes between teleological explanations which are and those which are not genuinely "purposive", where what he means by "purposive" is precisely that feature of "intentionality" which I referred to above (Taylor 1964: 54-71). The difference between the two is that a purposive or intentional explanation explains by citing something about the agent's own view of the situation in which he acts rather than merely by citing some goal or end which the behaviour as a matter of fact brings or has a tendency to bring about. Intentional explanation fixes the behaviour as action rather than mere movement which brings about a certain end; it implies that the behaviour is directed toward a certain goal or end, an end which the agent has in mind and plans to achieve by the behaviour.

Having ventured a brief exposition of what mechanistic and intentional explanations basically are, the issue of whether they are compatible or not can now be directly addressed.

The question is whether there can be both a mechanistic and an intentional explanation of a single instance of behaviour. Compatibilists claim that there can be; incompatibilists claim the opposite.

Now it might appear that compatibilists have common sense quite clearly on their side, for it seems obvious that for any ordinary instance of intentional action there is a mechanistic story to be told which can go at least some if not all the way towards explaining that behaviour. Suppose that I raise my arm in order to signal the waiter and that this is a full-fledged intentional action. Now my arm's going up certainly is a physical event which evidently is brought about by antecedent physical events - nerve impulses and muscular contractions most probably - and citing these would constitute a mechanistic explanation of why my arm went up. But the availability of this mechanistic explanation of why my arm went up can hardly be thought to be incompatible with it nevertheless being true that I raised my arm because I wanted to signal the waiter, i.e. with there being an intentional explanation of my behaviour. To suppose otherwise would be to suppose that I could raise my arm in a vacuum as it were, directly, without the usual physiological processes such as nerve impulses and muscular contractions occurring as well, and this we know to be just plainly false. So it just cannot be true that an intentional explanation of a piece of behaviour excludes there also being a mechanistic explanation of the physical events of which that behaviour is composed. At least some or other kind of explanatory co-existence is therefore indicated, and any incompatibilist will have to concede as much.

But does this sort of compatibility demolish the incompatibilist's entire argument?

I think not.

All that is required is to point out that the compatibility in this instance is but just a case of that general condition of compatibility mentioned in Section 1.2, where part of what makes it possible for one set of laws and the corresponding explanations to have application is precisely that another set of laws and corresponding explanations - perhaps more basic and wider in scope than the former - also have application and partially account for the fact that the former laws and explanations hold. (Remember that Kepler's Laws hold because Newton's Laws hold.) In this instance, part of what makes it possible for me as an embodied agent to raise my arm and hence for there to be an intentional explanation of that action is just that the right mechanistic laws (of neurophysiology, physiology and biology) also apply to me (or at least my body). If they did not, then it just would not be possible for me to act in any effective way. As Swinburne argues:

It is for this reason that the motion of a human hand is often explicable both by personal and by scientific explanations. The motion of my hand may be fully explained by goings-on in the nerves and muscles of my arm, and by physiological laws. It may also be fully explained by my bringing it about, having the intention and power so to do. Yet in this case the causes and reasons cited in each explanation provide a partial explanation of the occurrence and operation of the causes and reasons cited in the other. The goings-on in my nerves and muscles are ... brought about unintentionally by my bringing about the motion of my hand intentionally. Also, the operation of physiological laws provides part of the explanation of my having the power to move my hand - only because nervous discharges are propagated as they are, am I able to move my hand (1979:45-46).

What the incompatibilist wants to insist on, however, is that even though the physical events which comprise an instance of action are susceptible to a full mechanistic explanation, it cannot be thought that such a mechanistic explanation can be extended backwards so as to include among the causes of the behaviour physical events occurring 'outside' the agent.

Put more crudely, if I raise my arm for some or other reason, then that physical event cannot be completely explained by antecedent physical events stretching back indefinitely to physical events occurring 'outside' or 'beyond' me, perhaps even to a time before I was born. If it could, then there would be a complete mechanistic explanation of the motions of my body which would make nonsense of the claim that I did anything or that my intentions and purposings were somehow causally relevant to the motions of my body which in fact came about.

Antony Flew (a former compatibilist it might be noted!) puts the point this way:

... nothing has been said to preclude the possibility - which is surely being ever more abundantly realised - of finding physical causes for most of what goes on when a moving occurs. The only thing which it is contradictory to suggest ... is that there might be a complete explanation in terms of physical causes, all in their turn fully so caused, of the moving itself. Thus, when I move the little finger of my right hand there may be - there surely are - ongoings within the right arm which contingently necessitate appropriate contractions in the muscles of that finger. But what there certainly cannot be, if the moving is truly a moving and not a motion, is an unbroken chain of sufficient physical causes stretching back indefinitely. It would therefore seem that the central nervous system must either be or contain an apparatus of which the total input does not contingently necessitate every element of total output (1978:117).



What Flew is arguing here is simply that if an instance of behaviour is a genuine instance of action rather than mere movement, then there cannot be an explanation of that behaviour which totally by-passes the agent whose action it is. But in order to secure an explanation which terminates with the agent - <sup>26</sup> in order for the behaviour to be action rather than just part of the broader sequence of events - there cannot be a full mechanistic explanation (i.e. a deterministic explanation) of every physical event involved in that action where every such event can itself be fully explained mechanistically and so on backwards to the point where the explanation cites causes 'outside' and independent of the agent. Flew's guarded comment about the nervous system either being or containing "an apparatus of which the total input does not contingently necessitate every element of total output" can simply be taken as a commitment to physical indeterminism as a necessary condition for the very <sup>27</sup> possibility of action.

What the incompatibilist wants to exclude, then, as being incompatible with the reality of action, is that type of explanation which traces the causes of an action to some physical events occurring 'outside' the agent whose action it is, where such 'external' causes can fully account for the

26. Perhaps one need not insist that explanation stop with the agent, but only that it somehow include the agent. The 'agent' might just be an element in a causal chain and not a prime-mover unmoved as Chisholm (1982) would have it. See Richard Taylor (1966:126-129) for arguments to the effect that agent-causation is compatible with determinism. The point in any case still stands.
27. Flew has endorsed this interpretation in a personal communication.

behaviour in question. Let us call this type of explanation a complete explanation and say that an event is completely explained if it is fully explained by some preceding event, or series of events, which has its origin in an object or event external to the object undergoing the event being explained.

What is intended here can perhaps best be illustrated by a simple example: we have two meters A and B which are to register a reading, A being activated by wind-pressure, let us imagine, and B by an inbuilt Geiger-counter which registers the particle decay of some radio-active substance located within it. Assume that modern physics is quite correct about the physical indeterminacy of quantum events within meter B. Now the reading on meter A can be completely explained by the mechanistic design of the meter and physical events occurring outside it which cause it to register just what it does in a perfectly predictable manner. But the reading on meter B cannot be explained in this way. There are no physical events occurring outside the meter which cause it to behave as it does, and when we look inside the meter for an explanation we come to a point where there just are no further physical events to be cited as causes. At most we can have a partial explanation in terms of probability distributions. The point that requires emphasis is that even though the actual fluctuations of the meter-needle can be fully explained mechanistically - perhaps in terms of electrical input and resistance - the chain of causes cannot be traced back indefinitely to some cause outside the system, and it is the physical indeterminism of (parts of) the system which secures this fact.

In much the same way, the incompatibilist will concede, it might well be the case that the motion of my arm, when I intentionally raise it, can be fully explained mechanistically and also intentionally and hence that the two sorts of explanations are to that extent compatible. But what he denies is that such a mechanistic explanation could ever be a complete explanation, for it is this, he insists, which would be incompatible with the motion of my arm being an intentional action. So while it is true, as compatibilists claim it is, that there can be both a mechanistic and an intentional explanation of a single instance of behaviour, it is the incompatibilist's contention that this can only be true on the condition that the mechanistic explanation is not a complete explanation of the behaviour in question.

In Section 1.1 an account was offered of what seem to be the prima facie grounds for explanatory incompatibilism, and the basic argument developed there suggested that mechanistic explanations tend to displace intentional explanations. That basic argument can now be modified to the claim that mechanistic explanations tend to displace intentional explanations insofar as they are complete explanations of the phenomena in question.

The reasoning behind this argument can now be summarised as follows:

(1) intentional explanations have actions rather than mere movements as their proper subjects, and actions either include or have among their causal antecedents some or

other mental components or volitional state of the agent so acting, whether these be his desires, beliefs, willings, intendings, goals and reasons or, as agent-causalists argue, just the agent himself.

(2) when behaviour comes about independently of an agent's bringing it about or independently of an agent's mental or volitional states, then that behaviour can hardly be thought of as action at all, but is rather just a sequence of movements which have very little if anything to do with the agent, other than possibly being motions of parts of his body.

(3) but if a complete mechanistic explanation of behaviour is possible, then that is enough to show that neither an agent nor any other of his mental or volitional states were in fact causally efficacious in bringing that behaviour about, and so by (2) above, the behaviour cannot be thought of as action.

(4) therefore there cannot be both an intentional and a complete mechanistic explanation of a single instance of intentional action.

And that, I take it, constitutes the basic argument for  
28  
explanatory incompatibilism.

Now there are, at least insofar as I can discern, three general strategies which a compatibilist might employ in trying to avoid the conclusion of the argument above, each strategy corresponding roughly to a denial of each of three

28. Norman Malcolm (1968) offers a version of this argument in his "The Conceivability of Mechanism".

premises. Each strategy will be discussed in detail in the following three chapters.

The first strategy - what I refer to as the "Double-Language Strategy" - amounts to the denial that actions have causes at all, the language of action and its explanation being so different from that of physical events and their explanation that the two can hardly be rivals, as Flew so aptly puts it (1978:95), for the occupation of the same area of logical space.

The second strategy, which is discussed under the heading "Instrumentalism" consists in denying that intentional explanations entail any commitment to regarding the behaviour in question as being produced by or involving something "mental". Instead it regards the entire "mentalist" language of action, agency and intentionality as being merely an explanatory device which is ontologically neutral with respect to the actual mechanisms underlying behaviour.

The third strategy, what I refer to as "Physicalism", involves the claim that the desires, beliefs, intentions or reasons which are said to be the causes of action are themselves physical events which can thus feature in a mechanistic explanaton of action. Thus a mechanistic explanation of behaviour need not be incompatible with an intentional explanation of that behaviour because the physical states cited as causes are identical with the intentional states picked out in the intentional explanation.

Each of these strategies seem to me to be defective, and in the pages that follow I hope to show why.

## CHAPTER TWO

### THE DOUBLE-LANGUAGE STRATEGY

My aim in this chapter is to discuss, and then reject, the first of the three strategies a compatibilist might employ in trying to reconcile mechanistic and intentional explanations of behaviour.

It will be recalled that the first premise of the basic argument for explanatory incompatibilism entails a commitment to some or other version of the causal theory of action. It states that those of our bodily movements which are or are part of actions are caused by antecedent mental events such as willings, intendings, desires and beliefs and so forth, or else just by the agent himself.<sup>1</sup> It is this which makes mechanistic and intentional explanations potential rivals, for if it can be shown that a given instance of behaviour can be completely explained mechanistically, then this would leave no room, so the argument goes, for there also being causally efficacious mental antecedents of the behaviour - that is unless the two sorts of descriptions and explanations are systematically related as the Identity Theory claims they are.<sup>2</sup>

1. The causal theory of action is usually understood as the Davidsonian position, i.e. the position which holds that those of our bodily movements which are actions are caused by the desires and beliefs which rationalise them. I use the expression so as to include other causal theories such as the "volitionist" theory of Pritchard which claims that acts of will are the causes of our voluntary and intentional behaviour, as well as the agent-causalist theories of Chisholm (1976) and Richard Taylor (1966).
2. See for instance Ryle 1980; Melden 1961; Kenny 1963.

So one obvious way of avoiding the alleged conflict is to deny that intentional explanations are causal explanations at all. And this is indeed what proponents of the double-<sup>3</sup> language theory do deny. Rather than regarding the reasons why an action was performed as referring to the mental causes of the action, these philosophers argue that the language-games of offering reasons and citing causes are too different to be thought of as rivals for the occupation of the same area of logical space. And because they are so different, they cannot but be compatible. David Scarrow (1981:13) sums up the position quite neatly:

This was the position that interpreted reasons for behaviour as non-causal and the causes of behaviour as non-rational. It was a double-aspect theory: when behaviour was explained in terms of reasons that behaviour was understood as action; when behaviour was explained in terms of causes that behaviour was explained as physical, bodily movement. There could be no question of how a physical change could bring about human action, or of how a human reason could bring about a physical change. Anyone who raised this old mind-body conundrum was committing a category mistake: he was illegitimately homogenising two different languages. It is as if one were to ask how it is possible to checkmate a king with a piece of ivory.

The classic statement of the position is to be found in Melden's Free Action (1961:184):

Where we are concerned with causal explanations, with events of which the happenings in question are effects in accordance with some law of causality, to that extent we are not concerned with human action at all but, at best, with bodily movements or happenings; and where we are concerned with explanations of human actions, there causal factors and causal laws ... are wholly irrelevant to the understanding we seek. The reason is simple, namely, the radically different logical characteristics of the two bodies of discourse we employ in these different cases - the different concepts which are applicable to these different orders of inquiry.

3. See for instance Ryle 1980; Melden 1961; Kenny 1963.

The view Melden is expressing here can, of course, be traced back to Ryle's classic study The Concept of Mind. Like Melden, Ryle was concerned to demolish the Cartesian myth of the ghost in the machine according to which mental phenomena exert pushes and pulls on our bodies - our pineal glands - thereby raising our bodily motions to the status of free, rational and moral actions. Ryle's demolition of the Cartesian myth, as well as his alternative picture - logical or philosophical behaviourism, is well enough known so as not to require rehearsal here. What interests us is how Ryle sought to accommodate complete mechanistic explanations for our actions with the reality of human, rational action. In his famous chapter on "The Will" in The Concept of Mind Ryle addresses the issue directly. Says Ryle:

I have spoken of Mechanism as a bogey. The fear that theoretically minded persons have felt lest everything should turn out to be explicable by mechanical laws is a baseless fear. And it is baseless not because the contingency which they dread happens not to be impending, but because it makes no sense to speak of such a contingency. Physicists may one day have found the answers to all physical questions, but not all questions are physical questions (1980:74).

The truth of the matter, Ryle goes on to suggest, is that mechanistic explanations of behaviour can exist alongside intentional explanations of behaviour so long as it is realised that these are descriptions and explanations at two totally different levels, so different that there in fact can be no competition between them. As an analogy Ryle offers us the example of a game of chess. Each piece on the board, he says, obeys the laws of physics, but this is not to suggest that each piece is thereby constricted to follow a pre-ordained set of moves which precludes the players playing according to the rules of the game and playing intelligently.



or stupidly. If one wanted to know why a particular piece on the board moved from position A to position B, then an explanation in terms of the laws of motion and physiology might very well be satisfactory. But if one wanted to know why the repositioning of a knight resulted in checkmate for the opposing king, then an explanation in terms of the laws of physics would be wholly inappropriate. To explain this one would need to refer to the rules of the game and the various strategies which are permitted within the legal boundaries. And this sort of explanation would not be in competition with the explanation in terms of the laws of physics. Rather they are explanations of different aspects of one and the same process, and as Ryle says, "what the illustration is meant to bring out is the fact that there is no contradiction in saying that one and the same process, such as the move of the bishop, is in accordance with two principles of completely different types and such that neither is 'reducible' to the other, though one of them presupposes the other" (1980:76).

Applying all this to the case of human action and its explanation, the moral of the story ought to be clear: it is only if one thinks that human actions are bodily movements which have a certain 'mental' causes that one is bound to think that there is some tension between mechanistic and intentional explanations. For this would mean that the physiological causes of our movements must in some way be incomplete in order to leave room for the mental causes. But the mistake here is to suppose that it is the business of action-explanations to provide causes of the actions, and this, Ryle argues, is to misunderstand the logic of the language of action. Described as physical movements, our

behaviour consists in a complex series of physiological occurrences the explanation of which would consist in citing the relevant physiological causes and laws. But described as actions, our behaviour consists in a complex set of interactions which can only be explained by placing them in the context of the rules, conventions and institutions of the society in which they occur, and this is not a matter of finding causes - mental or physical - at all.

Two questions naturally suggest themselves at this point. Firstly, why is it supposed that the explanation of the action in terms of reasons is not to be understood as causal explanation, and secondly, if citing reasons is not causal explanation, then what is it?

In reply to the first question, proponents of the double-language theory normally bring two sorts of considerations to bear - the so-called Logical Connection Argument in order to show that reasons cannot be causes, and then the Contextualist theory of action in order to show that there cannot be causal laws which can state necessary and sufficient conditions for human actions.

Firstly the Logical Connection Argument. Appeal is popularly made to Hume's requirement that causes be logically distinct from their effects, but since all the likely candidates for causes of actions - motives, desires, intentions, beliefs, etc. - are all logically connected with the actions of which they are meant to be the causes, they cannot be distinct in the required way. Therefore reasons cannot be causes. This

argument appears in various guises in Ryle, Melden, Kenny,<sup>4</sup>  
Whiteley and Richard Taylor.

Melden, for example, writes as follows:

Let the interior events which we call "the act of volition" be mental or physical (which it is makes no difference at all), it must be logically distinct from the alleged effect - this surely is one lesson we can derive from a reading of Hume's discussion of causation. Yet nothing can be an act of volition that is not logically connected with that which is willed - the act of willing is intelligible only as the act of willing whatever it is that is willed (1961:53).

Later on, when discussing the relation between "wanting" and "desiring" and action, Melden says:

If the relation were causal, the wanting to do would be, indeed it must be, describable independently of any reference to the doing. But it is logically essential to the wanting that it is the wanting to do something of the required sort with the thing one has. Hence the relation between the wanting to do and the doing cannot be a causal one (1961:128).

The same sorts of considerations can be applied to all the candidates for causes of our actions, and in each case the conclusion would be the same: all these items are conceptually related to the doings of which they are meant to be the causes, and since the relation is conceptual it cannot be causal, for causes must be logically distinct from their effects. Thus when I raise my arm because I want to signal the waiter, the 'because' here is not to be understood as implying that my wanting to signal the waiter is a distinct event which stands in the relation of cause to the arm's

4. Ryle 1980:87-88; Melden 1961:52-55, 83-89, 115, 191; Kenny 1963:80; Whiteley 1968:106-108; R. Taylor 1966: 222-223.

going up, for if this were the case, then there must be some way of referring to and describing the 'wanting' independently of the action of which it is meant to be the cause. But this surely cannot be done, for to say that I want to signal the waiter is just to say that I am disposed to act in such a way as is appropriate to bring about the desired result, and in this context - the restaurant say - raising one's arm just is the action of signalling the waiter.<sup>5</sup> There are not two actions, that of raising my arm and also the action of signalling the waiter, but only the one action which is describable both as raising my arm and as signalling the waiter.<sup>6</sup> So to say that I want to signal the waiter is just to say that I was disposed to raise my arm because in this context raising my arm is the very same action as signalling the waiter. But in that case the 'wanting' is conceptually connected with the doing, and so cannot be a cause of the doing. So far from citing some cause of my arm going up by saying that I did it because I wanted to signal the waiter, as though my wanting to signal the waiter is a distinct event which can be discovered independently of any doing, the action of raising my arm is explained by describing further what sort of an action it was, what its context was, what the pattern is into which it fits given the rules, conventions and institutions of the society in which it occurs.

A related consideration which double-language theorists bring to bear in order to show that the explanation of action does not involve the citing of mental or physiological causes nor

5. Melden:152-153, 169.

6. Anscombe 1979:11; Davidson 1980A:14, 1980C:56-61.

the application of causal laws, is that which draws its inspiration from the so-called contextualist theory of action.<sup>7</sup> This is the view that actions are constituted as the actions they are not by any specifiable physical or physiological features, but rather by the norms, conventions<sup>8</sup> and social institutions of the society in which they occur. The reason for this is that an indefinite number of different physical movements can all satisfy the description of being the performance of one and the same action type, and a number of identical bodily movements can all count as totally different action types depending on the context within which they occurred. Thus the action type of making a payment can be realised in a variety of action tokens, from signing a cheque to handing over coins, and the act of raising an arm may count as the action type of signalling the waiter in one situation and also as the action type of delivering an insult in another situation or society.

P.F. Strawson has argued that "if every case of someone's telling a lie were an instance of one physically specifiable class of sets of physical movements (and vice versa), and every case of someone's jilting his girlfriend were another such case (and vice versa), then the reign of law in physical movements would mean that lying and jilting could be deterministically explained. There would, as far as lying and jilting were concerned, be physical laws of human action as well as physical laws of physical movement. But there is

7. Melden 1961:179-198; R.S. Peters 1958:12-15; Anthony Kenny 1973:93-97.

8. For a fuller account of the view see Shaffer 1968:94.

no question of such correlations ever being established". The upshot of all this is meant to be that since from a physiological description only we could never infer exactly which action has been performed, an explanation of behaviour in terms of physiological causes and laws would never amount to the explanation of the action. Or as Melden (1961:200) says: "absolutely nothing about any matter of human conduct follows logically from any account of the physiological conditions of bodily movements".

Human actions, then, are explained not by citing causes and causal laws, but by the complex business of relating them to the beliefs, desires, goals and intentions of the agents who perform them in the contexts in which they are performed. And this, the double-language theory maintains, is a business of exhibiting the logical - not causal - relations which may or may not hold between these various items/factors. It is this which makes them susceptible, unlike mere physiological happenings, to rational and moral appraisal. Described as bodily movements, human behaviour is a series of physiological occurrences "for which physiological occurrences would appear to be sufficient causal conditions";<sup>10</sup> described as actions which agents perform, the behaviour can only be explained by reference to the rules, conventions and institutions of the community within which they occur. So far from mechanistic and intentional explanations being rivals, they are too far apart and too different to be in

9. Pears 1963:66.

10. Melden 1961:200.

competition. It is only the illegitimate application of the causal model to human action and its explanation which can make them appear to be rivals. But once this "fatal blunder" has been spotted and eradicated, it should be obvious that there is no question of incompatibility between these two sorts of explanations. Once it has been appreciated that the explanation of human action constitutes an autonomous language-game within which the rules and logic of the game do not permit the language of cause and effect - as these concepts are employed in the physical sciences - then it can be seen how both sorts of language-game, that of action and that of bodily movement, can both have their proper application and place without in any way encroaching upon one another.

The question, then, of whether a complete mechanistic explanation of human behaviour is possible becomes irrelevant to its characterisation and explanation as human action which can be rationally or morally appraised. As Kenny says":

Even one hundred per cent predictability at the level of physiology need not by itself involve any increase in predictability at the human level. For physiological laws will enable us to predict physiological effects from physiological causes: and we shall need in addition at least translation-rules from the language of physiology into the language of human behaviour (1973:94). 12

And this possibility, the argument concludes, cannot be realised.

So even unlimited success in the physical explanation of human behaviour would leave unimpaired the fact that the

11. Melden 1961:202.

12. See also Melden 1961:214-215.

physical motions thus explained can also be viewed and appraised as the intelligent or stupid, moral or vicious, intentional actions of free and rational human agents.

Having been exposed to a view like this, one might be forgiven for simply sighing despairingly and saying "If only it were that easy". For the first problem with this double-language theory, indeed the central problem, is that it just seems to conflict head-on with one of our most cherished intuitions, namely that our reasons, desires, intentions and goals, in short our mental lives, do make a difference to what in fact happens in the physical world. And it is hard to see what sort of a difference this could possibly be if it is not a causal difference. By keeping reasons and causes so far apart, as this view does, our mental lives begin to look more like shadowy epiphenomena than the causally efficacious realities we intuitively take them to be. As Scarrow (1981:13) argues:

An arsonist's desire to see a building in flames and his beliefs about how to accomplish this end can bring about physical destruction. If his desires and beliefs had been different the building he destroyed would still be standing. And a counterfactual of this sort signifies causality: it signifies that the reasons and beliefs played a causal role in bringing about a physical eventuality. Socrates' claim that his physical presence in prison must be explained in terms of his moral convictions rather than in terms which altogether omit his beliefs, principles, and decisions still stands as a challenge to any version of the double-aspect theory.

According to the double-language theory, when we say that John raised his arm because he wanted to signal the waiter, the 'because' here is to be understood entirely non-causally. It is rather to be understood as generating further descriptions of the action, descriptions which relate



action to the desires, beliefs and goals of the agent in the light of which the behaviour can be seen as fitting into a rational pattern. The explanation of action, on this view, is more like justification than the normal sort of causal explanation encountered in the physical sciences. Davidson (1980A:10) admits that actions are explained in this way: "when we explain an action, by giving the reason, we do redescribe the action; redescribing the action gives the action a place in a pattern and in this way the action is explained". But Davidson denies "that, because placing the action in a larger context explains it, therefore we now understand the sort of explanation involved. Talk of patterns and contexts does not answer the question of how reasons explain actions, since the relevant pattern or context contains both reason and action"(1980A:11).

Davidson has also pointed out that even though intentional explanations do involve an element of justification which the more usual causal explanations do not, one cannot infer from this that explanation by reasons cannot also be causal explanation with the feature of justification being just one distinguishing mark (1980A:9). The problem, Davidson points out, is that there is a difference between providing a reason as justification for what one did and providing the reason on which one acted - "a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action because he had the reason"(1980A:9). But how is the double-language theorist who denies that reasons are or can be causes to endorse the force of this 'because'?

How is he to endorse the difference between this 'because' and the because of mere justification unless it is by recognising that in the one way case the reasons are causally operative and in the other not?

13

I have already mentioned that double-language theorists appeal to the logical connections between reasons and actions in order to deny that reasons are or can be causes. The fallacy in this move, however, has been conclusively demonstrated by Davidson, especially in his "Actions, Reasons and Causes" (1980A) which set out to reaffirm the Socratic position that reasons make a causal difference to the way the world is. The basic mistake, Davidson argues, is to think that events themselves rather than the descriptions of the events can be logically related. If this were true, then it would be impossible for the cause of B to cause B, because the description of the one event contains a reference to the other. But the logical relation here is purely grammatical - it depends on how the events are described rather than on what events are described. Once this has been appreciated, it can easily be seen that 'the cause of B' is but one description of the event picked out by another of its descriptions, viz. 'A', and so even though there may be descriptions of the events which logically relate them, the events themselves are not logically related and hence there is no obstacle to their being related as cause to effect.

13. Davidson's remark that "if, as Melden claims, causal explanations are 'wholly irrelevant to the understanding we seek' of human action, then we are without an analysis of the 'because' in 'He did it because...', where we go on to name a reason" (1980A:11) is surely pertinent here.

Thus even though it might very well be true that we cannot refer to or describe a reason apart from the action it explains, this would be no obstacle to the reason and action being causally related. Causal relations, so the slogan runs, are extensional, they hold between events however they are described.<sup>14</sup> So if reasons are not to be regarded as causes, then it will have to be for some reason other than their conceptual connection with the actions they explain.

The next criticism concerns Melden's claim that explanations of bodily movements and explanations of actions are so far apart that "absolutely nothing about any matter of human conduct follows logically from any account of the physiological conditions of bodily movement" (1961:200). This surely is an exaggeration. G. Warnock has pointed out that while it may be true that from a purely physiological account of some happening we may never be able to infer which action had been performed, it might nevertheless be possible to infer<sup>15</sup> that some particular actions had not been performed. As Antony Flew comments: "It is not hard, for example, to think up physiological descriptions from which we could infer that the person so described had not sat down in a chair, bowled an over overarm, or committed adultery" (1979:113).

What is probably true about Melden's claim is that there could not be physiological conditions both necessary and sufficient for the performance of some action type, such as delivering an insult. The reason for this is that the action type can be realised in an indefinite

14. Davidson 1980F:243.

15. Pears 1963:77.

number of physical ways, or, as R.S. Peters has put it, "we can never specify an action exhaustively in terms of movements of the body or within the body".<sup>16</sup> But what one is not entitled to infer from this is that there are no physiological conditions both necessary and sufficient for the performance of each and every particular action token, for what is true of action types need not be true of action tokens. The point is that while there may never be physical laws of human actions as such - no physical laws specifying what would count as the delivery of an insult for example - one cannot conclude from this that the delivery of an insult on a particular occasion cannot be accounted for in terms of physical causes and laws. C.H. Whiteley (1973:65) puts it this way:

The concept "disaster" can occur in no physical law; for what makes a happening a disaster is its relation to human desires and values, with which physics is not concerned. But this does not prevent such a disaster as the Lisbon earthquake from being completely accounted for as the effect of physical causes operating according to physical laws ... Nor, from the fact that human actions as such cannot be causally accounted for in physical terms, are we entitled to conclude that they cannot be causally accounted for at all. The argument assumes, and does not show, that all causes are physical and that explanations in terms of the apprehension of reasons are not causal.

Davidson (1980A:13) raises a slightly different but related objection:

Melden asks the causal theorist to find an event that is common and peculiar to all cases where a man intentionally raises his arm, and this, it must be admitted, cannot be produced. But then neither can a common and unique cause of bridge failures, plane crashes, or plate breakings be produced.

The problem with the double-language view, as Flew for

16. R.S.Peters 1958:12.

has pointed out, is that if both the language-game of explaining bodily movements and the language-game of explaining actions are to have application to this one world, then we should expect that the two languages would intersect each other at some or other point. <sup>18</sup> And so we know that they do. For human actions often are or include items of bodily movements. When I raise my arm in order to signal the waiter, to use this now well-worn example, then it follows that my arm goes up. But this familiar sort of bodily movement needs to be distinguished from the case where though it is true that my arm goes up, I do not raise it - it happens to me, a reflex motion, say, of some or other sort. But how, according to the double-language theory, is this distinction supposed to be made? <sup>19</sup> For it has to be admitted that the genuine case of the raising of an arm is no more and no less a bodily movement than the mere rising of an arm, but if the one is an action and the other not then the difference between the two, even (or especially one could say) at the physiological level, must surely lie in the way in which these two items of bodily movements are generated, i.e. in their causal origins. And in that case we are led back to some sort of causal theory of action - we are led back to saying that some items of bodily movement are actions in virtue of having certain characteristic causal antecedents which mere bodily movements do not have.

17. Flew 1975:114. But see also Richman 1969:48; J.R. Lucas 1970:20.

18. Kant's solution to this problem was to deny that we in fact have only one world. See Boyle 1976:100-121.

19. Some writers on the topic of action think that this is the central problem of the entire philosophy of action. See Frankfurt 1978:157.

The double-language theorist, it seems, in order to be consistent, would have to say that we distinguish between those bodily movements which are or are part of actions and those which are mere happenings in some way other than appealing to causes. But what could this other way be? Melden says that it is through "training" that we learn to see other people's behaviour as action, as behaviour which is to be understood in terms of the distinctively human concepts of action, reason, motive and purpose. He says:

But how is it possible for us to see a person raise his arm, to see a bodily movement as an action? Well, how is it possible for us to read a printed page, to see, not curiously shaped black marks on a white background, but the sentences that lie before us? Here the answer is simply 'Training'.

Melden, no double, has a point here. But the analogy he suggests is surely misconceived. For none of us have ever thought or believed that a printed page in itself signifies anything. None of us have ever thought that the squiggles on the page signify anything independently of what, by the conventions we have, we take them to signify. But the case with action is different. For here there does seem to be some fact of the matter about what a person is doing which is independent of how it is perceived or regarded by others. When a person is raising his arm, then irrespective of how this is seen by others, whether as a signalling or as an insult or whatever, there surely is one description of what he is doing which has a status untouched by all other perceptions and descriptions, and that is the agent's own description of

what he is doing. My point is that whether something is an action or not depends on what the agent is doing and not just on how the bodily motions are seen by others. And this is unlike the printed page which depends for being a piece of meaningful writing precisely on being understood as such in virtue of the conventions of a particular society.<sup>21</sup> In itself it is merely a collection of squiggles.

The same point can be pursued somewhat differently. If we ask ourselves the question "Why should we regard a piece of bodily movement as an action?", then, if the double-language theory were true, there would be no plausible answer. For if there is nothing 'internal' to the behaviour which makes it an action, if it were entirely a question of choosing to describe and explain it as such, then we could conceivably describe any piece of behaviour as action regardless of whether there are intentions, reasons or goals present at all, and this seems decidedly odd.<sup>22</sup> For we can readily imagine a new breed of sophisticated robots which resemble genuine human persons in all respects except that they have no mentality whatsoever. Are we to 'see' their behaviour as action also? Are we to suppose that they are capable of action just because their behaviour can be explained and understood within the language-game of human action and its explanation? If this were the case, then the distinction between action and mere bodily movement, of which Melden

21. Charles Taylor (1985B:193) makes a roughly similar point in the connection with the problem of attributing actions to machines.

22. This is, of course, the instrumentalist approach to the question favoured by Dennett which will be discussed in the next chapter.

makes much, would appear to become obscure indeed, for then there would be no reason - apart, perhaps, from a reduced capacity to successfully predict and explain - not to ascribe reasons and intentions to even the simplest of mechanisms and explain their behaviour according to the pattern of human action. The conclusion ought to be that if there is a difference between actions and mere bodily movements, then the difference cannot consist simply in the way we happen to explain these two sorts of occurrences, but must consist in something 'internal' to the occurrences themselves, and it is hard, perhaps impossible, to see what this could be if not some difference in the causal antecedents - described mentally or physically - of the two sorts of bodily movements.

Ryle, it will be remembered, attempted to show that "Mechanism" is a "mere bogey", that "not only is there plenty of room for purpose where everything is governed by mechanical laws, but there would be no place for purpose if things were not so governed" (1980:70). He says:

The favourite model to which the fancied mechanistic world is assimilated is that of billiard balls imparting their motion to one another by impact. Yet a game of billiards provides one of the simplest examples of a course of events for the description of which mechanical terms are necessary without being sufficient. Certainly, from accurate knowledge of the weight, shape, elasticity and movements of the balls, the constitution of the table and the

23. This connects with our view that we believe in other minds, not by virtue of some inference from our own case (the argument from analogy), but because it is the best explanatory hypothesis we have at this stage. See Churchland 1986:67-72 as well as Dennett 1973:164.



conditions of the atmosphere, it is in principle possible, in accordance with known laws, to deduce from a momentary state of the balls what will be their later state. But it does not follow from this that the course of the game is predictable in accordance with those laws alone. A scientific forecaster, who was ignorant of the rules and tactics of the game and of the skill and plans of the players, could predict, perhaps, from the beginning of a single stroke, the positions in which the balls will come to rest before the next stroke is made; but he could predict no further. The player himself may be able to foresee with modest probability the sort of break that he will make, for he knows, perhaps, the best tactics to apply to situations like this and he knows a good deal about his own skill, endurance, patience, keenness and intentions.

But this seems entirely odd. For if everything in the universe, including every item of human bodily movement, is governed by mechanistic laws according to which it is in principle possible to make accurate predictions, then how is it possible to deny that the course of the game of billiards can also be so predicted? For here, as everywhere, every move of every muscle, every stroke of every cue, every movement of every ball is subject to these mechanistic laws, and the course of events on, in and around the billiard table is but the inevitable consequence of physical causes operating according to these laws. It is interesting to note, as Skillen (1984:517) points out, that "Ryle's discussion of the physics of billiards ends with the cue, at the other end of which is more or less 'skill, endurance, patience, keenness and intentions'.... So despite what he says elsewhere Ryle denies .... the Principle of Physical Determinism, since he denies that the physical location of a ball is in principle predictable in the light of physical knowledge".

The trouble is, of course, to make the "skill, endurance,

patience, keenness and intentions" relevant to what goes on on the table. If they are not relevant, then we do not have much of a game, at least not in any ordinary understanding of the term. We might as well regard the motion of the planets as a game. But if they are relevant, then the only way they can get purchase on what goes on on the table is by being causally operative factors, in which case we are led back, once again to some or other version of the causal theory of action.

In conclusion then, I submit, the view has to be rejected that mechanistic and intentional explanations can be made compatible by banning causality altogether from the arena of human action. Apart from the weakness of the arguments for this move, the view ends up obscuring the distinction between actions and mere events which it was its prime purpose to bolster. The difficulty is, I have argued, is that the language-game of action and the language-game of mere events do overlap - actions often are or at least include items of bodily movements, and since these presumably can in principle be completely explained mechanistically, it becomes difficult to see why they, or some of them at least, should be regarded as actions at all. It was admitted, of course,

24

24. Identity theorists who subscribe to the causal theory of action at least have an answer to this: actions just are those of our bodily movements brought about by our reasons and intentions, etc., and these are the very same events as those picked out in the mechanistic explanation of our behaviour.

that we could - and probably would - continue to regard these mechanistic occurrences as actions and hence choose to explain them as such, but this, I submitted, is contrary to what double-language theorists in any case set out to show. For it was hoped that by showing that the notion of causality is inapplicable to human actions, then, as Richman puts it, "we would have rescued human freedom (conceptually) at just the point - that of human action - which is essential for

26

morality". But far from showing that our status as free and rational agents has nothing to fear from mechanism, the position that emerges is that our intentions and desires are irrelevant to the motions of our bodies which in fact come about. Our freedom and agency, on this view, becomes little more than linguistic - it exists only insofar as we choose to explain our behaviour in the language-game of action rather than mere movement. We are free and rational, it would seem, to the same extent that a plant is when, undergoing the process of photo-tropism, we explain its behaviour not by reference to biological causes and laws, but by reference to its desire to acquire more sunlight and its belief that by moving a branch upwards to the left it will achieve this end.

But this, I think, is no freedom or agency at all. An incompatibilist can still insist that since the behaviour in question can be completely explained mechanistically, there is no reason to regard it as action in the first place. One

25. P.F.Strawson has argued that not only will it be irrational for us to do so, but also that we probably never could abandon our ordinary, intentional outlook on human behaviour. See Strawson 1980:1-25.

26. Richman 1969:42.

condition, at the very least, for something being an agent or for its behaviour being action, surely, must be that its behaviour is somehow produced or caused by its desires, intentions, goals and reasons.

## CHAPTER THREE

### INSTRUMENTALISM

#### 3.1 Intentional Explanations as an Explanatory Device

In the previous chapter I considered and then rejected the view that mechanistic and intentional explanations could be reconciled by keeping them in totally different categories, bodily movements occurring in the realm of causality but actions not. The fact is that these two language-games do overlap, and any plausible account of their relation will have to accommodate this fact.<sup>1</sup> Our actions for the most part are or at least include physical movements, and if these occur within the ordinary realm of causality, then presumably so do our actions, or at least parts of them. The problem, for the compatibilist, is to account for this fact. What he has to account for is the possibility that all our bodily movements can be completely explained mechanistically and at the same time also be the intentional actions of rational agents.

There are, at least insofar as I can discern, only two alternatives: either our intentions, desires, reasons and beliefs etc. are themselves physical events which can thus feature in a complete mechanistic account of our behaviour, as proponents of the Identity Theory claim. Or else there just are no such things as actions, agents, desires, reasons

1. "Melden, for example, says that actions are often identical with bodily movements, and that bodily movements have causes; yet he denies that the causes are causes of the actions. This is, I think, a contradiction." (Davidson 1980A:18). See also Melden 1961:74.

and beliefs which require to be accommodated within a complete mechanistic account of human behaviour, as eliminative materialists claim. Instead, these concepts are to be regarded as an explanatory device in the absence of a complete mechanistic understanding of behaviour. My aim in this chapter is to discuss this latter alternative.

Explanatory incompatibilism, to recall, can simply be formulated as the thesis that where and whenever a complete mechanistic account of a piece of behaviour is available, then that behaviour cannot be the intentional action of a rational agent.

Stated like this however, the view lends itself to a type of objection, what I have called the Instrumentalist objection, which is advanced by, among others, Daniel Dennett.

The objection, very briefly, is this: whether a piece of behaviour is to count as an instance of intentional action is not an ontological issue concerning the 'mechanisms' whereby the behaviour in fact comes about, but is rather a question of whether it makes good predictive and explanatory sense to view the behaviour from within an intentional framework. The mistake of the incompatibilists, according to this view, is their failure to see that our usage of intentional

2. This is not intended to be a discussion of eliminative materialism as such, but only a discussion of how it is thought that this strategy can reconcile mechanistic and intentional explanations of behaviour. The discussion focuses exclusively on the writings of one exponent, namely Daniel Dennett, in the hope that the views expressed will be fairly representative.
3. For a similar point of view and related arguments see Rorty 1965; P. Churchland 1986 and P.S. Churchland 1986.

explanations is ontologically neutral with respect to the way in which behaviour is in fact 'produced' within a particular system. Since the intentional explanation of a piece of behaviour leaves it entirely open as to the actual mechanisms by which the behaviour comes about (and this is in any case a matter for science to settle), any underlying mechanistic account of the behaviour, if there is one at all, cannot be ruled out or be incompatible with the intentional account.

The sort of argument put forward here is strongly reminiscent of those arguments for modal compatibilism which purport to show that the only coherent sense to the notion of free will is that which is ontologically neutral with respect to either determinism or indeterminism. The argument here is usually that the criteria whereby we pick out paradigmatic cases of free action are entirely mute as regards the actual causal histories of the actions involved, but concern rather the question whether the agent could or could not have done otherwise under the circumstances. This view is generally supported by appeal to the now famous 'conditional' analysis of the 'could' of the previous sentence, whereby to say that an agent could have done otherwise is just to say that he would have done otherwise had he so chosen or wanted. And this concept of free action is compatible with determinism.

4. In his The Explanation of Behaviour Charles Taylor seemed to argue that the intentional explanation of behaviour ruled out there being an underlying mechanistic account. But in Taylor 1970 and Taylor 1985A he rejects all a priori arguments like that of Norman Malcolm's (1968) to establish the point and settles for the moderate position of regarding this as an empirical issue to be settled by us finding an account of human behaviour which is more basic than the intentional account.
5. See for example Smart 1963:122-126. For a modern and sophisticated version of this argument see Dennett 1984.

In much the same way, the argument goes, we pick out paradigmatic intentional actions not by examining whether the behaviour comes about mechanistically or not, but rather by seeing whether it makes sense to speak of reasons for action in the context. If it does, then the behaviour can plausibly be viewed as intentional action, irrespective of whether there in fact is an underlying mechanistic structure to the behaviour or not.

Dennett develops his concept of an intentional system with its attendant intentional stance with the explicit aim of forming, as he says, "a bridge connecting the intentional domain ... to the non-intentional domain of the physical sciences"(1979:22). An intentional system, says Dennett, is a system whose behaviour can be successfully "predicted and explained by relying on ascriptions to the system of beliefs and desires (and hopes, fears, intentions, hunches, ...)"(1979:3). In viewing a system as an intentional system one adopts what Dennett calls the intentional stance toward that system, and the hallmark of the intentional stance is the presupposition it makes regarding the rationality of the system in question. One assumes that the system is rational, and it is only after discovering lapses in or limits to its rationality that one adopts the mechanistic stance toward it.<sup>6</sup> Now there certainly are numerous instances of behaviour where we adopt the intentional stance toward a system even though we know of perfectly good mechanistic explanations of that

6. Dennett 1973:164. Dennett distinguishes two mechanistic stances: the design stance and the physical stance. The difference between them need not concern us here and I will use 'mechanistic stance' to refer only to the latter.



behaviour, however cumbersome or superfluous these in fact may be (Dennett 1973:164). Such is the case when we adopt the intentional stance toward a chess-playing computer. We assume that the computer will make rational moves given its assessment of the situation, and it is only when we detect limitations in its design or breakdowns in its function that we adopt the mechanistic stance, as when we call in the IBM repair-man to reconnect some circuits. Similarly with guided missiles. We adopt the intentional stance toward them when we talk of their goals, their pursuit of and homing in on the target, etc. We assume that the system will make rational responses and adjustments according to its perception of the situation, and it is only when we suspect a malfunction in its mechanical working out of rational responses that we adopt one of the mechanistic stances.

In adopting the intentional stance toward a system one does not make any assumptions as to whether the system really is conscious, or really has beliefs, desires or intentions. One merely ascribes these intentional states to a system as an explanatory device without regard for any ontological commitments they may imply. Dennett (1979:7) says:

Lingering doubts about whether the chess-playing computer really has beliefs and desires are misplaced; for the definition of intentional systems I have given does not say that intentional systems really have beliefs and desires, but that one can explain and predict their behaviour by ascribing beliefs and desires to them, and whether one calls what one ascribes to the computer beliefs or belief-analogues or information complexes or intentional whatnots makes no difference to the nature of the calculation one makes on the basis of the ascriptions.

And elsewhere (1973:164) he says:

The success of the [intentional] stance is of course a matter settled pragmatically, without reference to

whether the object really has beliefs, desires, intentions and so forth, so whether or not any computer can be conscious, or have thoughts or desires, some computers undeniably are intentional systems, for they are systems whose behaviour can be predicted, and most efficiently predicted, by adopting the intentional stance toward them.

What then of the relation between the two stances, or the relation between mechanistic and intentional explanations?

Dennett argues that what gives rise to the feeling that there is some sort of antagonism between them is the absence from the mechanistic stance of the presupposition of rationality which we make when adopting the intentional stance. But, he goes on to argue, the crucial point to be realised is that an absence of a presupposition of rationality is not the same as a presupposition of non-rationality.

... from any particular mechanistic explanation of a bit of behaviour it would not follow that that particular bit of behaviour was or was not a rational response to the environmental conditions at the time, for the mere fact that the response had to follow, given its causal antecedents, casts no more doubt on its rationality than the fact that the computer had to answer '108' [(when calculating the product of 18 and 6)] casts doubt on the arithmetical correctness of its answer (1973:173).

Dennett seems to think that it is the relative simplicity of the examples that are normally used in this debate that enhances the illusion that mechanistic and intentional explanations cannot co-exist, but he goes on to point out that "this is a case where the philosopher's preference for simple examples leads him astray, for of course any simple mechanistic explanation of a bit of behaviour will disqualify it for plausible intentional characterisation, make it a mere happening and not an action, but we cannot generalise from

simple examples to complex, for it is precisely the simplicity of the examples which ground the crucial conclusion"(1973:173). The only thing to be inferred from the fact that a system's behaviour can be explained mechanistically is not that the system is not rational, but rather that the system must in some extended sense be tropistic - it can never be designed so as to ensure a rational response to every possible set of conditions that it might encounter. This means that the system will always be imperfectly rational because, as Dennett argues, the notion of a perfectly rational system is the notion of an unrealisable physical system.<sup>7</sup>

Dennett here has in mind those philosophers who have argued that behaviour which is brought about mechanistically cannot by that fact be rational, as witness for instance Alastair MacIntyre's statement that "if a man's behaviour is rational then it cannot be determined by the state of his glands or any other antecedent causal factor".<sup>8</sup> There are also the arguments of writers like J.B.S.Haldane and C.S.Lewis to the effect that determinism (or mechanism or naturalism) is self-defeating because if everything (including our actions and our beliefs) is the outcome of blind and non-rational mechanistic causes, then so too is our belief in mechanism and hence cannot be thought of as a rational belief.<sup>9</sup>

The mistake in all these arguments, Dennett points out, is

7. Dennett (1973:172) offers a thought-experiment involving the re-designing of wasps to establish this point, but the notion of 'rationality' he uses here, so I shall argue, is ambiguous and does not establish the point he in fact is trying to make.
8. MacIntyre 1957:248.
9. See Lewis 1964; R.E.Ewin 1968; Flew 1975:92, 229; Smart 1963:126-128; Popper and Eccles 1983:75-81.

that they all fail to register the fact that the absence of a presupposition of rationality is not the same as a presupposition of non-rationality, and that there is no reason for thinking that a piece of behaviour which is mechanistically produced cannot also be the rational response of that system or organism to its situation at the time. Or as Ayer points out in his The Concept of a Person (1964:267), there is no reason for supposing that a system cannot operate both causally (mechanistically) and according to the laws of logic (rationally). To be sure, it will be pointed out, assessing the rationality of some bit of behaviour or belief is a task entirely independent of any investigation into the actual causes of the behaviour or belief, so that any upshot of this latter investigation will be entirely mute insofar as the former inquiry is concerned.<sup>10</sup>

J.J.C.Smart (1963:127) writes as follows:

[If] determinism is true, then all arguments are caused. From this it follows that some causally determined arguments must be good arguments. For among these causally determined arguments must be some which have passed the independent tests for rationality. It follows that the argument that naturalism is self-defeating is not valid.

Smart does not say what the independent tests for rationality are, but presumably he has in mind the simple checking whether or not a conclusion follows logically from some set of premises. Thus if a computer is given the information that 'All men are mortal' and 'Socrates is a man', and if it then goes on to conclude that 'Socrates is mortal', then

10. This indeed seems to be the general response of all those writers who think that the argument that determinism is self-defeating is invalid. See for instance Anscombe 1981A and also footnote 9 above.

presumably the computer's response is rational even though it is causally determined to answer just as it in fact does.

What Dennett's arguments above have made abundantly clear, it seems, is that there is a perfectly straightforward sense in which a purely mechanistic system can be said to 'embody' or 'manifest' what we normally refer to as the intentional states that are associated with intentional explanations. In this sense, then, mechanistic and intentional explanations are indeed compatible. We need to see, according to this view, that intentional explanations are merely an explanatory device which neither entail nor presuppose any restrictions on the possibility of there also being a mechanistic account of behaviour. And while adopting the intentional stance toward a system presupposes that the system is (roughly) rational, one is not to suppose that the absence of this presupposition from the mechanistic stance entails that the system in question is therefore non-rational, for, as Dennett has argued, nothing about whether a piece of behaviour is a rational response or not follows from a mechanistic account of its behaviour.

But it is at just this point, I want to argue, that an ambiguity concerning the way the term 'rationality' is used begins to manifest itself, and, so I shall argue, once this ambiguity has been cleared up, then a whole lot of what Dennett has to say on the relation between mechanistic and intentional explanations loses its force and leaves us with substantial grounds for still claiming that there is some sort of incompatibility between the two.

### 3.2 Designed and Enacted Rationality

The ambiguity in the notion of 'rationality' as it has been used in the discussion above comes out clearly when we ask of some particular action or belief whether it is rational or not. For what is this supposed to mean?

Consider the case of belief first.

When we ask of some particular belief whether it is rational or not, then it seems that we could be asking one of two possible questions. Firstly, we could be asking whether it is in general rational for someone to hold that belief, and, secondly, we could be asking whether a particular person is rational in holding that belief, and the two sorts of questions are very different.

Take for instance the belief that the world is spherical. Now it does seem as though it is in general rational for people to hold this belief, and the criteria of rationality in this instance are what we might call objective, relating to publicly ascertainable data such as the overwhelming evidence for the truth of the belief.

But when we ask whether a particular person is rational in holding this belief, then the criteria of rationality are no longer objective in the same sense as before, but relate rather to the person's entire belief-set in the light of which this particular belief may or may not be rational or logically justifiable. The person might have in her belief-set other beliefs among which the belief that there are no spherical objects or the belief that only square objects can support life may feature. In such a case, even though the belief that the world is spherical may be rational (because it is true or because it is the best account of the facts),

the person herself is not rational in holding that belief.

Davidson has called attention to this distinction in his work on the paradoxes of irrationality by the following sorts of remarks:

Here we must make an obvious distinction between having a reason to be a believer in a certain proposition, and having evidence in the light of which it is reasonable to think the proposition true. (Sentences of the form 'Charles has a reason to believe that p' are ambiguous with respect to this distinction.) A reason of the first sort is evaluative: it provides a motive for acting in such a way as to promote having a belief. A reason of the second kind is cognitive: it consists in evidence one has for the truth of a proposition (1985:143).

And also:

Much that is called irrational does not make for paradox. Many might hold that it is irrational, given the dangers, discomforts ... for any person to attempt to climb Mt Everest without oxygen (or even with it). But there is no puzzle in explaining the attempt if it is undertaken by someone who has assembled all the facts he can, given full consideration to all his desires, ambitions and attitudes, and has acted in the light of his knowledge and values. Perhaps it is in some sense irrational to believe in astrology, flying saucers, or witches, but such beliefs may have standard explanations if they are based on what their holders take to be the evidence. It is sensible to try to square the circle if you don't know it can't be done (1982:290).

The distinction is that between what may be called externalist criteria of rationality as opposed to internalist criteria of rationality. The former appeals to public and objective factors, while the latter appeals to what goes on within an agent's mind, what his other beliefs, desires and attitudes are, and in this sense is subjective rather than objective.

The upshot of all this is that when a person is rational in holding some belief, then this is entirely different from saying merely that the belief is rational. For in the former

case but not the latter there is a requirement that the person be internally justified (in terms of her other beliefs) in holding that belief, and that is to imply that the person holds the belief not merely because the belief is capable of being logically justified by external criteria, but also that the person holds the belief because she recognises it as being logically consequent upon her other beliefs and this, in part, accounts for her present belief. (The point here is similar to Davidson's point that a person can have a reason for an action, and perform the action, and yet this not be the reason on which the person acted.) It is her seeing the logical connections that justifies her belief and makes her rational in holding it, and this demands that she be a conscious agent.

I now want to consider this distinction in relation to the explanation of actions, because it is here that a blurring of the distinction can lead to the blurring of a distinction between different types of intentional explanations, and this can enhance the plausibility of Dennett's compatibilist claims.

Dennett has persuasively argued that in adopting the intentional stance toward a system there is no requirement that the system really be conscious or really have beliefs, desires and intentions. This indeed seems to be true. But, I want to show, the rationality which this stance presupposes is only of the former 'externalist' kind whereas it is rather the second 'internalist' kind of rationality that is crucial in our ordinary intentional explanations and, indeed, in our ordinary understanding of what actions are. Dennett shows



that the former externalist-type intentional explanations are compatible with mechanistic explanations, but this goes nowhere near showing that the latter internalist-type are also.

The distinction I wish to draw corresponds to the externalist/internalist distinction drawn above and manifests itself in the two different types of rationality we may suppose to be present in a system when adopting the intentional stance toward it. The first type, what I will refer to as designed-rationality, presupposes no commitment as to whether a system really is conscious or has beliefs, etc., while the second, enacted-rationality, it seems, necessarily does. Just as one cannot think of a person as being rational in holding a particular belief without that person actually having perceived the logical connections between his other beliefs upon which the rationality of his present belief is dependent, so too one cannot suppose that a system is rational in the enacted-rationality sense without supposing that the system is conscious and has beliefs, desires and intentions.

As an example of designed-rationality, consider the way in which I (automatically) withdraw my hand when it accidentally comes into contact with a hot stove. This is not a simple knee-jerk reflex as when a doctor taps my knee, but is rather something which I do as opposed to something which merely happens to me. It is not an intentional action, but is an action all the same - it is something which I can bring under my conscious control, or refrain from doing (with a great deal of effort no doubt), or acquire greater dexterity or

the laborious framing of the right intentions and the bearing in mind of hundreds of little details eventually pays off when the 'rational' connections become 'hardwired' into the biological or neural system, and you are no longer required to think or reason about what you are doing but can manage on designed-rationality alone.

12

I am not suggesting that these sorts of actions are simply like those pieces of behaviour which have some sort of biological drive as their source, but only that they are not the prime examples of rationality in action. They are rather habitual actions which are precisely such as to liberate us from the need to forever think and reason about what we are doing. We so to speak 're-design' ourselves by intentionally inculcating these sorts of habits, and though this 're-designing' requires that we be agents capable of operating on ourselves, the performance of these habitual actions does not manifest the richness and complexity of the rationality involved in full-fledged intentional action. There is no conscious decision to depress the clutch every time the gears are engaged, nor a conscious decision to accelerate at just the right point during this procedure. One just does this, and the reasons for doing it - and there are objectively good reasons - are not the reasons on which one is presently acting, for though one could state them if asked, they have as it were become incorporated, submerged, in one's character or nature. They are

12. See Charles Taylor 1970:76. See also Matson 1976:146 for a similar set of examples making the same point.

objectively present in the nature of the person so acting rather than subjectively and consciously present in his mind when he acts.

Enacted-rationality differs from designed-rationality in that it presupposes that a system in fact is conscious, because here the criteria of rationality are those I have identified as being 'internal' to the system.<sup>13</sup> When behaviour manifests enacted-rationality, then the behaviour comes about not merely because of certain antecedent states which can be described in such a way that the behaviour appears to be a rational response of the system at the time, but rather because of certain antecedent intentional states which rationalise the behaviour from the agent's own point of view. In this case, the agent does what she does because she has seen that it is the rational thing to do given her desires, values and beliefs, and it is rational precisely because of this perception of the logical relations between intentional states and not merely because there objectively are these logical relations.

For example, let us suppose that I find myself in a restaurant and it so happens that I want to signal the waiter. Suppose also that I believe that by raising my arm I

13. The distinction drawn here between designed and enacted-rationality has a rough parallel with the distinction Taylor draws between those teleological explanations which are merely teleological and those which are also intentional, including as an essential feature the agent's own description and perception of what it is doing (Charles Taylor 1964:54-71). Taylor seems to think that only the latter can qualify behaviour as action. See also Boden 1978:39-52.

can signal the waiter. Suppose further that I do in fact raise my arm. Now we would normally be inclined to say that this action of raising my arm is objectively rational given the desires and beliefs that I do have. But the trouble is that this is not enough to make me rational in my action, because it might just be the case that I am unaware of the logical connection between the contents of my desire and belief. It might be the case that I just somehow fail to connect up these two intentional states. So in order for me to be rational in raising my arm something more is required than there just being a logical connection between my various intentional states, and this something is precisely that I see the logical connection between the desire and the belief and this seeing accounts for why I did what I did. <sup>14</sup>

And it is in precisely this way that a reason for the action can become my reason for performing it. And this, it seems, requires some sort of mentality or consciousness.

And it is in this sense that a computer (at least not of the sort we are presently familiar with) can never be rational in doing whatever it is that it does. Our regard for the apparent rationality of the computer lies not in its logical manipulation of concepts and symbols, but rather in its mechanical manipulation of physical tokens which we regard as standing for symbols and concepts and as being logically related. <sup>15</sup>

14. See A. Shutte 1984:484 for a fuller discussion of this point.

15. For a similar point see Dretske 1985:24-27.

### 3.3 Two Sorts of Intentional Explanations

From the discussion above it should be clear enough that the sort of rationality Dennett has in mind when he asks us to regard a computer as an intentional system is the sort of rationality I have described as designed-rationality. It is the sort of rationality that does not require that the system in question be a conscious agent, but only that it be capable of being described in such a way that its behaviour can be seen to be rational in the light of intentional states we ascribe to it.

And Dennett has, I think, shown that there is no incompatibility between regarding behaviour as rational in this sense and there also being a mechanistic explanation of the system's behaviour. For being rational in this sense is merely being a system which can be described as operating in accordance with logical principles. There is no requirement here that the system itself perceive the logical connections between its various intentional states.

Thus it does seem plausible to suppose that there indeed is a mechanistic explanation of the automatic withdrawal of my hand from a hot stove. We can suppose that there is a fairly complex neurophysiological and biological story to be told which can explain this event, and this in no way impugns our conviction that the behaviour for all that is rational and can also be explained in terms of various desires and beliefs which I may be said to have.

But it is not at all clear that Dennett's argument shows that behaviour which is rational in the enacted-rationality sense can be explained both intentionally and mechanistically. For

here, so I have argued, behaviour comes about not merely in accordance with logical principles, but partly at least because the agent herself has seen something rational in acting the way she does given just the intentional states she recognises herself as having. The logical connection between what the agent wants and how she sees herself as getting it given her present situation is mediated through the agent's present awareness or state of consciousness and not through some biological mechanism which links up appropriate responses to stimuli in a rigid and deterministic fashion - as in the withdrawal of my hand from a hot stove.

Why this is a problem for Dennett's account of the relation between mechanistic and intentional explanations is just that the only sort of intentional explanation he seems to allow for is that which is entirely compatible with the absence of any mentality whatsoever in the system whose behaviour is being explained. Now while it is surely right to insist that we often do employ intentional explanations of this sort, namely the sort which assumes only designed-rationality to be present in a system, it nevertheless seems clear that this is not the only sort of intentional explanation that there is. For we often perform actions, all our intentional actions, which involve enacted-rationality, and in the explanation of these from the intentional stance we do assume that the system in question in fact is a conscious agent which does really have desires and beliefs irrespective of how these are realised in the physical hardware of the system if at all. If there are to be intentional actions, then it would seem that there also has to be mentality.

That Dennett is dealing with the externalist or designed conception of rationality becomes even more apparent when he rejects the view that a physical system can be a perfectly rational system (1973:172). He writes:

What conclusion should be drawn from this about human behaviour? That human beings, as finite mechanical systems, are not rational after all? Or that the demonstrable rationality of man proves that there will always be an inviolable terra incognita, an infinite and non-mechanical mind beyond the grasp of physiologists and psychologists? It is hard to see what evidence could be adduced in support of the latter conclusion ... since for every awe-inspiring stroke of genius cited in its favour (the Einstein-Shakespeare gambit), there are a thousand evidences of lapses, foibles, bumbling and bullheadedness to suggest to the contrary that man is only imperfectly rational. Perfection is hard to prove, and nothing short of perfection sustains the argument.

Dennett is probably right about most of this, but the peculiar thing is that he seems to have misunderstood the type of rationality that has seemed in the eyes of many writers to be threatened by mechanism. For it is not, as has already been conceded, the objective or externalist type of rationality that is in question, but rather the internalist or enacted type of rationality that seems to be at odds with complete mechanistic explanations of our behaviour. The point is that even though what a person does may be objectively irrational in the sense that it is just silly or stupid or contrary to our common sense, the person nevertheless has to be in some sense rational in order for it to be true that she did it. Davidson, for one, has pointed out repeatedly that all intentional actions have something rational at their core, and the reason for this should now be clear. For in order for a bit of behaviour to count as an intentional action,

16. Davidson 1982:293.

there has to be a description of it, namely the agent's own account of what she was doing, under which her desires and beliefs rationalise the action. The rationality resides not in the act which was done, but in the mind of the agent who performed the action. She has to see something rational in her action given her desires and beliefs - it is not enough that there just be something rational from our point of view. When MacIntyre says that "if a man's behaviour is rational then it cannot be determined by the state of his glands or any other antecedent causal factor", then it is not the objective rationality of what the man does given certain desires and beliefs attributed to him that is in question, but rather the rationality of the man himself because in order for him to be rational he has to act, partly at least, because he sees that it is the logical thing to do and this seeing is a causal factor in explaining why he did it. And it is this internal rationality which is threatened by mechanism because the man's internal rationality is apparently shown to be causally irrelevant to the motions of his body which in fact come about. But this, it seems, would undermine the very possibility of intentional action, because it is this type of rationality, enacted-rationality, rather than the externalist or designed type of rationality which is crucial in our understanding of what intentional actions are.

17. See Shutte 1984. I am not suggesting here that logical relations cannot be explicated in terms of causal processes, for while I think that this is true and constitutes a problem for any materialist position, my concern here is just to show that Dennett does not seem to appreciate that our rationality differs in an important respect from the demonstrable rationality of computers and the like. See Madell 1986:160-164 for an account of the former issue.



Dennett is mistaken in thinking that the adoption of the intentional stance toward a system presupposes no commitment to regarding that system as a conscious agent. This would be true if the only intentional stance we had was the one which assumed only designed-rationality to be present in a system. But that does not seem to be the way it in fact is. For it still seems to be a legitimate question to ask of some piece of behaviour whether it was an intentional action or not, and this question cannot simply be answered by saying that we can regard the behaviour as though it was an intentional action by ascribing beliefs and desires to the system whose behaviour it is. For that is not enough to make something an intentional action. As Michael Levin has pointed out (although in a slightly different context):

The ease with which we predict the behaviour of a chess-playing computer by thinking of it as wanting and planning may indeed be evidence that the computer really wants and plans to win. But this sort of ascription of intensional states depends on a prior understanding of literal instances of wanting and planning. Indeed, it sometimes appears that the ultimate aim of the as-if strategy is to assimilate the ascription of intensional states of human beings to the as-if attitude we occasionally take to feedback systems. Apart from undercutting itself, such a proposal flies in the face of the fact that whether I fear snakes does not depend on whether observers can predict my behaviour if they assume I am. To say I fear snakes is not to say something about the most convenient way to think of me, but something about the way I actually am (1979:171).

Dennett's instrumentalism does seem to conflate the question of whether something in fact is an intentional action with the question of what would count as evidence for something being an intentional action. The first question is an issue concerning the identity conditions for something, the second is a question of establishing whether those conditions obtain or not. Now it does seem to be true that our usual evidence

for counting a piece of behaviour as an intentional action consists in seeing whether in the context it makes sense to speak of reasons for the behaviour and not in examining the causal ancestry of the physical motions involved. But this is a far cry from saying that being an intentional action just is a question of it making sense in the context to speak of reasons for the behaviour. Flew has argued that we should acknowledge our fallibility in this regard (1975:116), for there is always, he says, "a theoretical possibility of misidentifying a motion as a moving". But this should not tempt anyone to think that there is no fact of the matter as to whether something is an action or just a mere movement. And if some piece of behaviour in fact is an intentional action, then the bodily motions involved must, so I argued in the previous chapter, be caused at least in part by the intentional states which rationalise the action. The problem, once again, is that of finding room in a basically physicalistic ontology for the causal efficacy of these mental states. And Dennett's variety of compatibilism falls far short of the mark in showing how this is possible. For he seems to leave out of his account precisely those actions, namely full-fledged intentional actions, which are the obvious candidates for failure to be caught in the mechanist's net.

In conclusion, then, it would appear that Dennett's notion of an intentional system with its attendant intentional stance poses no threat to incompatibilist claims. For the incompatibilist can insist that while it may not be true that offering an intentional explanation of a bit of behaviour rules out there also being a mechanistic explanation of that

behaviour, this is so only on the assumption that intentional explanations are of a uniform sort. But if what has been argued above is sound, then this assumption can no longer be sustained, for we have identified two types of intentional explanation which employ different conceptions of rationality, and while the one type is indeed compatible with mechanism, nothing has been said to show that the other is also. We can, it would seem, adopt an instrumentalist attitude toward our ascription of intentional states to a system if the intentional stance we are dealing with assumes only designed-rationality to be present in the system. But, I have argued, intentional actions manifest enacted-rationality, and enacted-rationality requires that a system in fact be a conscious agent. So to show that this kind of intentional explanation is compatible with mechanistic explanations, one would have to argue, it would seem, that the intentional states ascribed to a system, even a conscious and rational system (in the enacted-rationality sense), are themselves physical states which could thus be referred to rather than excluded in a mechanistic explanation. One would have to show that an agent's conscious perception of logical relations is itself a series of physical events. Compatibilism, it seems, cannot be established without facing up to this challenge.

## CHAPTER FOUR

### PHYSICALISM

The previous two chapters were attempts to dismiss those arguments for explanatory compatibilism which do not rely on the suggestion that desires, beliefs, reasons and intentions - in short, all the usual candidates for causes of our actions - are themselves events which can thus feature in a mechanistic explanation of behaviour. Dennett's instrumentalism, of course, has some or other version of physicalism as its ultimate aim, but he does not - at least not in the works cited - set about establishing explanatory compatibilism by explicitly arguing that desires and beliefs are themselves physical events. Instead he attempts to show that these concepts are merely part of an explanatory hypothesis we adopt towards the behaviour of certain systems, and the explanatory hypothesis in no way requires that a system really has desires or beliefs or even be conscious. This strategy fails however, and the reason it fails is that it seems to ignore features which are essential to and partly constitutive of intentional action, namely genuine mentality and enacted-rationality. So the ultimate problem for the compatibilist remains that of showing how the causal efficacy of the mental can be compatible with there also being complete mechanistic explanations of our behaviour.

In his "The Conceivability of Mechanism" Norman Malcolm presents an argument for explanatory incompatibilism which closely parallels the argument I set out at the end of

Chapter One. Malcolm's argument, roughly, goes like this: if we say that somebody climbed a ladder because he wanted to retrieve his hat from the roof, then this explanation is understood to have certain counterfactual implications - for example, that had he not wanted to retrieve his hat then he would not have climbed the ladder. And this is usually understood as implying that the man's intention to retrieve his hat was causally relevant to his motions up the ladder. But now, should there be a complete mechanistic explanation of his movements up the ladder, then this counterfactual would not hold at all, for the mechanistic explanation would give all the sufficient neurophysiological conditions of the man's movements and his intention or desire would not be among these. Thus, Malcolm concludes, if there is a complete mechanistic explanation of the man's movements up the ladder, then it would not be true that he climbed the ladder because he had the relevant desire or intention. Says Malcolm:

Given the antecedent neurological states of his bodily system together with general laws correlating these states with the contractions of muscles and movements of limbs, he would have moved as he did regardless of his desire or intention. If every movement of his was completely accounted for by his antecedent neurophysiological states (his 'programming', then it was not true that these movements occurred because he wanted or intended to get his hat (1968:52).

Malcolm's argument has, of course, been objected to by many,<sup>1</sup> and the crux of these objections seems to be that Malcolm does not take seriously enough the suggestion that human behaviour can be described and explained from within two

1. See for instance Goldman 1970:157-169; Larry Wright: 130-132; Long 1979:130-140.

different levels of explanation, namely the intentional and the mechanistic, assuming as he does that these would always be rivals. But the problem is that in order for two explanations (or two causes) to be rivals, they would have to be, as Flew says, rivals for the occupation of the same area of logical space. That is, they must be competitors at the same level of description and explanation. But this would involve us regarding intentions or desires and beliefs or agents as such, rather than their neurological correlates, as likely causes of physical and neurological events, and this would be extremely weird indeed. It would be resorting to an undiluted Cartesian dualism according to which mental events as such can effect physical changes. Douglas Long expresses the point quite succinctly:

A sufficient neural explanation of a movement excludes as causes of that movement all and only those other events which are competing causes. By a competing cause I mean one which can intelligibly be supposed to have produced that effect instead of its actual cause. For a neural account to occupy the explanatory place of the man's intentions, such intentions would have to be at least in the running as causes of bodily movements in a way that rivals physiological explanation. The two must be competitors at the same level. But the idea that an intention, as such, and not some physiological correlate of it, can cause a physical event like a motion of a hand is incoherent (1979:133).

In Chapter One I argued that there cannot be two unrelated or independent causal explanations of an event where these explanations are full explanations, citing necessary and sufficient conditions for the occurrence of the event. The reason was that if causes C1 and C2 are both necessary and sufficient for the occurrence of event E, then causes C3 and C4 cannot also be both necessary and sufficient for E because this would contradict the idea that C1 and C2 are necessary

and sufficient.

The problem, then, with the idea that mechanistic and intentional explanations are not rivals because they are explanations at two totally different levels is that of accounting for how these two levels of explanation are related, because related they have to be. If the man's intention to retrieve his hat is a causal factor in explaining his movements up the ladder, and if his movements up the ladder can also be completely explained by citing neural and muscular events sufficient for bringing about movements, then there has to be some relation between these two explanations if we are to avoid some sort of causal overdetermination or epiphenomenalism.<sup>2</sup> The question is how.

I now want to consider two general proposals for how these two sorts of explanations can be related in the hope of showing that neither of them are satisfactory. The first, the so-called Correlation thesis, postulates that each and every mental event is correlated with a specific physical event such that there are psycho-physical laws which state, for example, that any entity will be in a state of having the intention to retrieve a hat at time  $t$  if and only if it is in a certain physical (neural) state  $N$  at time  $t$ .<sup>3</sup> The second proposal, the Identity thesis, postulates that each and every

2. McGinn 1979:40.

3. Like the Identity thesis, the Correlation thesis comes in a number of guises. One may hold that the psycho-physical laws hold across the board such that every kind of mental event is correlated one-to-one with a specific kind of neural event, a view that is currently unpopular and probably false anyway given the so-called multiple realisability of mental states. Or one may hold that these laws are species-specific, or that the psycho-physical laws will be different for each individual. See Goldman 1970:162-164. These different positions, I think, are irrelevant to my discussion.

mental event is identical with some or other physical event such that there may or may not be (depending on which version of the Identity theory one opts for, namely the Type or the Token Identity theory) a systematic way of translating the mentalistic language into the language of physical events.

Let me deal with the Correlation thesis first.

According to this view, a person can only be in a certain intentional state if the person also is in a correlated neural state in the sense that being in that specific neural state is both necessary and sufficient for being in that intentional state. From this it would follow that a person could not have the intention to retrieve a hat if the correlated neural state did not occur as well, and hence there being a neurological account of a person's movements would not exclude there also being an intentional explanation of his behaviour. Long writes as follows:

(If) it is possible to accept this minimal physicalism, it is possible to suggest that had the man not wanted his hat, certain crucial physiological events would not have occurred which did in fact take place, and had they not taken place his legs and arms would not have moved on the ladder as they did. Hence, although the neurological account quite properly does not mention his intention, it is not accurate to say, as Malcolm does, that it explains the movements 'without regard to his intention' (1979:132).

Alvin Goldman seems to present a similar argument in his rejection of Malcolm's incompatibility thesis.<sup>4</sup> Goldman introduces the notion of what he calls "simultaneous nomic equivalents" and argues that if there is a law stating that

4. Goldman 1970:161-164.



any object O has property X at time t if and only if O has property Y at time t, then O's having X at time t is a simultaneous nomic equivalent<sup>5</sup> of O's having Y at t. He then goes on to argue that if two events C1 and C2 are simultaneous nomic equivalents, then if C1 is necessary and sufficient for a further event E, then, by the principle of the substitutivity of necessity, so also is event C2 necessary and sufficient for event E. And this would mean that if there are psycho-physical laws which could render intentional and neurological states as simultaneous nomic equivalents, then it would follow that a man's movements up a ladder can be caused by his intention to retrieve his hat and also by the neurological state which is its simultaneous nomic equivalent.

The problem with all of this, however, is that even though we may agree that, in the example above, C2 is in some sense necessary for E, this does not by itself show that C2 is causally necessary for E as was intended. Goldman wants to hold that there could be laws of nature such that a man's intention to retrieve his hat is a simultaneous nomic equivalent of a certain neural state N occurring in the man's brain. And he also wants to hold that the man's intention to retrieve his hat caused his movements up the ladder. But the problem is that causes are usually understood to not only precede their effects, but also to have direction.

If events A and B are causally related, then it cannot, on any usual understanding, be the case that A caused B and also

5. This is not Goldman's precise and exact formulation, but it accurately expresses what he says. See also William Robinson 1979:328.

that B caused A. Causal relationships are meant to be unidirectional. So if the man's intention caused his movements, then presumably we have a case of pure mental causation as in the old Cartesian picture and Goldman surely does not want to argue for that. He rather wants to say that the intention can be construed as a cause precisely because it is a simultaneous nomic equivalent of a neural event which causes the appropriate physical movements. But does this show that the intention is causally necessary for the man's movements? I do not see how it can. It will not do to say, as Goldman does, that:

the grounds for saying that C2 is necessary for E are precisely those presupposed in saying that C1 is necessary for E. C1 is said to be necessary for E because, given the laws of nature, the omission of C1 (at t) would have meant that the events at t would not be sufficient for the occurrence of E. Similarly, given the laws of nature, the omission of C2 (at t) would have meant that the events at t would not be sufficient for the occurrence of E. Given the laws of nature, the omission of C2(at t) would have necessitated the omission of C1 (at t). And the omission of C1 as we have seen, would have resulted in the non-occurrence of E (1970:162).

For suppose that there is a law of nature that all and only oranges have five pips, and suppose further that there is a law of nature such that only oranges can be used for treating some particular disease D, then it follows that an object can only have the property of being a cure for D if it also has the property of having five pips. An object's having the property of being able to cure D would on Goldman's criterion be a simultaneous nomic equivalent of an object's having five pips. But I do not think that any of us would be inclined to accept the view that this is enough to show that an object's having five pips is therefore a cause of a patient's being cured of D, nor would we be inclined to accept the view that

an object's having five pips is causally necessary to whatever curative effects it might have. Of course it would be true that unless an object had five pips then it could not be a cure for D, but the reason for this need not be that having five pips is causally necessary for the curative properties of oranges, but rather that unless an object had five pips then it could not possibly be an orange and there is something else about oranges - their chemical structure say - which plays a causal role in curing D.

To make the having of five pips a causally salient feature of an orange's curative properties we would need not merely a law correlating oranges with objects having five pips, but a law which specifies what it is about the having of five pips which is causally efficacious in curing the disease. And this, I take it, would require at least that some other feature of oranges, their chemical composition say, cannot completely account for the oranges's curative properties. And this would mean that even though an orange's curative properties are a simultaneous nomic equivalent of its having five pips, these two states of the orange are causal competitors at the same level. If the one completely accounts for the orange's curative properties, then there can be no law which shows that the other is causally<sup>6</sup> necessary as well. It might be necessary in some other sense of 'necessity', but it cannot be causally necessary and yet play no role in bringing about the desired effect.

6. This example was specifically designed to make the point abundantly clear, but for two less outlandish examples see W. Robinson 1979:329.

The same considerations apply in the case of having an intention to retrieve a hat and its alleged neurological correlate. That they are simultaneous nomic equivalents hardly shows that the intention is causally necessary to the man's movements up the ladder. His movements are completely accounted for by neurological causes and laws, and even though it may be true that these neurological causes could not occur unless the relevant intention occurred as well, this does not make the man's intention causally operative in bringing about the motions of his limbs up the ladder. The intention could just as well be an epiphenomenon, an idle by-product of the neural processes which, though nomically related to them, is completely determined by them and wholly impotent to causally affect the brain states or bring about any movements. The problem, as William Robinson has pointed out,<sup>7</sup> is that we understand, in principle at least, the physical laws and causation involved in brain states bringing about bodily movements, but it would seem that in addition to the laws correlating brain events with mental events we would need psycho-physical laws which show how mental events are causally connected with brain or other physical events in order to make them causally efficacious. Correlating them with brain events is not enough. Presumably neither Goldman nor Long would want anything to do with this kind of psycho-physical interactionism, but they can only avoid it at the cost of admitting that intentions (and the mental life as such) are mere epiphenomena after all. They might very well be correlated one-to-one with brain states, but their being

7. William Robinson 1979:328.

thus correlated makes them no more causally necessary to movements (actions) than movements of my shadow being correlated with my physical motions makes them - the shadows - causally necessary to the physical effects the motions of my body bring about.

It would seem then that the Correlation thesis fails to make intentions and the like causally efficacious in those of our movements which are our actions. It establishes a kind of explanatory compatibilism but only at the price of making our intentions causally redundant. It is, therefore, a kind of epiphenomenalism, distinct from the older variety in that it claims that it would not be true that the history of the world would have been the same had there been no mentality whatsoever, but an epiphenomenalism nevertheless in that it fails to show how mental events as such can be causal ingredients of our world. It is interesting, once again, to see just what would have to give in order to make this view compatible with genuine mental causality. As I argued above, there would have to be causal laws linking mental events with physical events where the direction of causality goes from the mental to the physical. But in order for this type of interactionism to be viable, then contra Long, intentions would have to be competitors with neural events at the same level in the sense that if movements are completely explained by the causal efficacy of prior neural events, then there could be no causal law which states that intentions are somehow also required to get the motions going. The only way that intentions could play a causal role is if the neural causes do not completely determine the movements which in fact come about but require the causal

effectiveness of intentions as well. And this, it will be recalled, is precisely what the incompatibilist claims: there cannot be a complete mechanistic explanation of a bit of behaviour where this behaviour is the intentional action of an agent, and the reason is just that there being such a complete mechanistic explanation of the behaviour is enough to show that no intentional states, no mental events, were causally efficacious in bringing the behaviour about. I conclude, then, that the Correlation thesis on its own presents no obstacle to the explanatory incompatibilist claims advanced, for instance, by Malcolm in the passage quoted above. It still remains to be seen how this kind of physicalism, when combined with mechanism, can get mentality into the causal picture.

One way of doing this, of course, is to claim that the causal effectiveness of mental states can be guaranteed if mental events are not merely correlated with neural events, but in fact are identical with them. In this way, it is believed, whatever causal effectiveness neural events have mental events will have too. This is the position apparently<sup>8</sup> favoured by most explanatory compatibilists.

Starting with his classic "Actions, Reasons and Causes", Donald Davidson, for instance, has developed a persuasive account of intentional action which not only gets intentions,

8. For example Davidson 1980A, 1980B; McGinn 1979; P. Smith 1981.

beliefs and desires back into the causal picture, but apparently obviates any mind-body hassles there might have been by showing that all these mental items or at least those which are causally connected with physical events, are themselves physical events which cause our motions along physiological lines we can readily comprehend. Davidson's view, Anomalous Monism, involves the following three seemingly inconsistent propositions:

10

- 1) Some mental events causally interact with physical events.
- 2) When events are related as cause and effect, then there is some strict deterministic law under which these events are subsumed.
- 3) There are not nor can there be psycho-physical laws which meet the requirements of (2).

The only way. Davidson argues, that these propositions can be rendered consistent is if mental events themselves are physical events which can thus feature in strict deterministic laws. The way to see this is to realise that events related as cause to effect do not have to be subsumable by a law which relates them under just any of their descriptions, but only that there be some description of them under which they are thus subsumable. Thus Davidson writes:

11

Suppose a hurricane, which is reported on page of

9. The paper was apparently written with the specific aim of reinstating the causal theory after its (temporary) demise in the hands of Ryle, Melden, Kenny and all those who accepted the now virtually defunct Double-Language Theory. See Scarrow 1981:13.

10. Davidson 1980B.

11. Davidson 1980B.

Tuesday's Times, causes a catastrophe, which is reported on page 13 of Wednesday's Tribune. Then the event reported on page 5 of Tuesday's Times caused the event reported on page 13 of Wednesday's Tribune. Should we look for a law relating events of these kinds? It is only slightly less ridiculous to look for a law relating hurricanes to catastrophes. The law needed to predict the catastrophe with precision would, of course, have no use for concepts like hurricane and catastrophe (1980A:17).

If reasons (or desires and beliefs, intentions, purposing, etc) are causes of our actions, then by (2) above, the so-called Principle of the Nomological Character of Causality,<sup>12</sup> there must be some description of these under which they can feature in a strict deterministic law, and since by (3)<sup>13</sup> above, the so-called Anomalism of the Mental, there "are no strict deterministic laws on the basis of which mental events can be predicted and explained" (1980B:208), these mental events must bear physical descriptions because these are the only descriptions of events which can feature in laws of the required sort. Thus mental events can enter the causal scene because they are identical with those physical events which cause our motions.

Does this strategy work?

Some recent writers on the topic have argued that it does not and the cause of the trouble they all trace back to one common source which we have already encountered, namely epiphenomenalism. They all argue that one cannot combine

12. Davidson 1980B:208.

13. Davidson 1980B:200. Davidson argues at length for these two principles in 1980B, and even though these can be challenged - see Anscombe 1981B for instance - it will not be my aim to do so here but rather to see what follows about mental causality once the argument has been accepted.



non-reductive monism with mechanism without this straightaway  
14  
leading to the epiphenomenalism of the mental.

Honderich (1982) asks us to consider the following scenario: a person places some green French pears on a scale at a market, and the pointer of the scale registers two pounds. We have a clear case of one event causing another. But, Honderich asks, what is it about the pears that causes the moving of the pointer? Now it surely is not in virtue of their being green and French that the pears cause the pointer to move, but rather it is in virtue of their weight that the pointer moves. The point is that not every property of an event is causally salient to its being a cause of some other event, but only those properties which figure in the causal law which explains how the two events are causally related. Honderich dubs this the Principle of the Nomological Character of Causally-Relevant Properties (1982:62).  
15

Applying this to the alleged causal interaction between mental and physical events as Davidson conceives of it, the question naturally arises as to which properties of a mental event (which is identical with a physical event) are the causally salient ones. And this question, Honderich argues, spells

14. See for example Hess 1981; Honderich 1982, 1983, 1984; H. Robinson 1982; A. Skillen 1984.

15. Similar examples are constructed by Robinson 1982:11 and by Skillen 1984:521. See also Charles Taylor 1970:56 for a clear statement of the view. The clearest and most convincing example is that of Honderich in his "Smith and the Champion of Mauve" where he challenges the Anomalous Monist to significantly distinguish himself from the man who claims that the "mauveness" of his slippers is causally relevant to their keeping his feet warm just because the mauve slippers are identical with those that do in fact keep them warm.

"bad news" for Anomalous Monism.

For if it is the mental properties of the event that are causally salient, then this would imply that there are after all causal laws which can relate the mental as such to the physical, and this would be contrary to the thesis of the Anomalism of the Mental whereby psycho-physical laws are excluded. And without this premise Davidson's argument for the identity of mental and physical events would in any case fall away.

But if it is only the physical properties of the mental event (which is identical with a physical event) that are the causally salient ones, then this would seem to render the mental properties of that event causally redundant - just as the greenness or the Frenchness of the pears are causally irrelevant to the movements of the pointer. And this would be a kind of epiphenomenalism. For it would amount to saying that the mental properties of an event might well have been absent without this in any way affecting the causal efficacy of the physical events with which that mental event is identical.<sup>16</sup>

16. In one of his replies to Honderich on this issue, Peter Smith (1984:84) points out that epiphenomenalism has traditionally been understood as the view that mental events do not cause physical events at all, and that Anomalous Monists do not subscribe to this view at all, believing as they do that mental events, being identical with physical events, do cause physical events after all. Honderich's reply (1984:88) is simply that this nevertheless is a kind of epiphenomenalism because it fails to give us the efficacy of the mental as such, for there is nothing in Anomalous Monism which makes an event's mental properties relevant to whatever physical events it may bring about. And with this, I think we have to agree.

It is for this reason that Howard Robinson (1982:10) says that Davidson's argument is "ingenious but sophisticated", because while he affirms the thesis that mental events sometimes cause physical events, Davidson "allows it a completely unnatural interpretation". Says Robinson:

He [(Davidson)] asserts ... that every mental event is also a physical event, although mental states and properties are not identical with any physical states or properties. Hence mental events can cause physical events, because they themselves are physical events. Thus, if my feeling a pain is identical with brain event B, and brain event B causes my hand to reach for my leg, then (the event of) my feeling a pain caused me to reach to my leg. However, the form of interactionism so generated is entirely empty. Indeed to call it 'interaction' seems disingenuous, for it remains the case that the mentality of the event played no part in bringing about its effect, and therefore is itself inert and epiphenomenal. Further, had the mental been absent no difference would have resulted (1982:11).

Robinson goes on to give an example - similar to Honderich's one of the pears - involving active and inert gases, and concludes that the "Davidsonian claim that mental events make a contribution, when combined with the admission that the mental state does not is fundamentally misleading. That this is how we understand causal interaction is shown by the fact that it is the lack of this type of contribution which justifies us in saying that certain gases are inert"(1982:11).

Is the charge of epiphenomalism correct?

An answer to this question is made extremely difficult by the fact that some of the issues involved, questions of identity and causality for instance, are still being hotly debated. There is the dispute between Davidson and Kim about what would constitute an adequate criterion for the individuation of events, Davidson subscribing to a causal theory and Kim

to a property exemplification account.

Davidson's causal theory allows him to hold that mental events can be identical with brain events even though mental properties are not identical with physical properties. Suppose that this is accepted, then it appears that Davidson would still face the problem of having these mental properties causally idle or epiphenomenal in whatever causal transactions that mental events as physical events might engage. This idea is reinforced by the thesis of the Anomalism of the Mental whereby nomic connections between the mental properties of an event and its physical properties are denied, claiming instead that whatever psycho-physical generalisations there may be will not be law-like. In one of his replies to Honderich, Peter Smith admits that "there are no nomological links between a mental event's mental properties and its causal physical properties in virtue of which it produces a certain upshot" (1984:85), and he seems to agree that this denial of the "unwanted psycho-physical laws" (83) is essential to Anomalous Monism. But how then can the mental as mental be brought into the causal picture? It would be useless to claim, as Smith appears to, that the efficacy of a mental event resides in its being identical with a physical event which we know and understand to have causal efficacy. For this slogan can be indefinitely repeated without it making the slightest difference to the fact that it is only the physical properties of an event which do all the causal work. In his illuminating "Smith and the Champion of Mauve", Honderich compares the situation to that of the man who claims that it would be a wonderful confusion to underrate the

mauvish efficacy of his slippers because his mauve slippers are identical with the slippers lined with Hibernian fleece which keep his feet really warm. Says Honderich:

The Anomalous Monist can be as wedded as he wants to the proposition that of course a mental event in his sense causes a physical event. By way of that truth he is no nearer getting mental efficacy than the Champion of Mauve is to getting mauvish efficacy by going on saying that it is the mauve slippers that keep him really warm.

I would conclude, then, that the charge of epiphenomenalism can indeed be made to stick, and that this is seriously damaging to the Anomalous Monist's case. For while we may indeed have our intentions (as physical events) as causal ingredients in our world, they are not causally efficacious in virtue of being an intention of this or that kind, but rather merely in virtue of being a neurological state of some or other sort. The causal efficacy (of the event) of my having the intention to raise my arm resides not in the intentional nature of the state that it is, but rather in its physical properties which are, according to the Anomalous Monist, not in any law-like or causal connection with the intentional or mental features of the event, namely its being an intention to raise my arm rather than to brush my teeth or phone Ronald Reagan. And this is, as Skillen (1984:520-524) says, to relegate the mental to the "realm of appearances", to accord "reasons a merely puppet status in respect of their causal power".

Relating all this back to the issue of explanatory

18. See also Swinburne 1986:100-102.

incompatibilism, it would seem then that the incompatibilist also has nothing to fear from this Anomalous Monist version of physicalism. And the reason is clear: if a piece of behaviour can be completely explained mechanistically, then the mental properties of the physical events which cause the behaviour cannot also be causally involved in bringing the behaviour about. They would be, strictly speaking, irrelevant to the motions which in fact do come about. And this is incompatible with regarding the behaviour as the action of an agent. Intentionality and agency, on this view, would be causally redundant features of the events thus described. And even though the mental properties of the events could in some sense be "piggybacking" on the physical events which cause the behaviour (the so-called <sup>19</sup> supervenience thesis), this would not make them causally effective features of the physical events. An incompatibilist can still insist - and with justification - that these sorts of events are not properly described as actions.

The supervenience thesis is meant to be the claim that objects or events cannot differ in their mental properties while agreeing in all their physical characteristics (Davidson 1980E:253). The physical features of an object are meant to determine its mental features but not the other way round. But I do not see what this claim can come to if it is also denied that there are nomological connections between the mental and the physical features of an object. Without nomological connections their relation would be accidental, and if accidental then there would be no reason

19. The phrase "piggybacking" is borrowed from Melchert 1986:272.

why my intention to phone Ronald Reagan is not a property of the physical events which cause my action of brushing my teeth. It is for this reason that Honderich says:

This Anomalous Monism, having denied psycho-physical connection, is so far indistinguishable from what we all hoped had been put to rest, which is to say mere psycho-physical parallelism. Will this be followed by a revival of Pre-Established Harmony? This Anomalous Monist's plight is illustrated by the fact that that would be better. It would offer some explanation of the fact in question (1984:89).

One way out of this is to attempt to reinstate psycho-physical laws, to find laws which link the having of a certain intention to being in a certain physical state. But this would be to deny the token-identity thesis and to repudiate Davidson's metaphysical argument for the identity of mental and physical events. The virtue of Davidson's argument was that it did not rely on science for finding invariable correlations between mental and physical events, as these were thought to be the only likely grounds for mooting an identity rather than mere invariable correlations

20. Denise Meyerson Taylor (1982) uses this lack of nomic connections to argue that the Anomalous Monist cannot believe that reasons causally explain actions. Her reason is that even though there are no laws relating hurricanes to the collapsing of bridges under those descriptions, citing hurricanes causally explains bridge collapsings because there nevertheless is a systematic way of translating from these descriptions to descriptions under which there would be a law which connects them. But this is not available in the case of reasons and actions precisely because the generalisations relating the mental to the physical are accidental, and hence statements "relating beliefs and desires to actions cannot be systematically related to statements expressed in a physicalist vocabulary". And this would mean that there cannot be causal explanation going on, but only coincidence. She urges instead that reasons do causally explain actions, but for the different reason that there are causal laws of the non-mechanistic teleological sort which can be employed.

between the two. His argument, instead, relied on denying that such correlations could be established. This, together with the common-sense belief that mental events do sometimes cause physical events and the metaphysical view that causal relations imply laws, yielded the conclusion that each and every mental event in causal interaction with physical events (and other mental events) is itself a physical event though there are no laws linking the mental and the physical under these descriptions. But if there are such laws, even undiscovered ones, then the argument fails and there is no reason, apart from arguments from the unity and simplicity of science and our having discovered at least some correlations to believe that mental events are physical events.<sup>22</sup>

Now I do not want to attempt a refutation of the identity thesis. It would suit my purpose quite adequately to allow that mental events are identical with brain events on the condition, only, that mental properties are not identical with physical properties. This, together with complete mechanism, is sufficient to show that mental properties are in fact epiphenomenal and this, I have argued, is just what the incompatibilist needs.<sup>23</sup> For it offends against our common-sense conviction that it is the mental as mental which plays a causal role in our actions. And I have already argued in this chapter that mere correlations will not be enough to give us the efficacy of the mental.

21. Davidson 1980B:212.

22. Smart 1963:11; Armstrong 1981:1-4.

23. Robinson (1982:10) states this consequence quite explicitly.



So the only way, it seems, that the specifically mental properties of events such as (the having of) intentions and the like can be incorporated into the causal fabric of the physical world, while complete mechanism is also true, would be for us simply to hold that mental events and properties are themselves physical events and properties, the sort of reductive materialism which Smart has advocated.<sup>24</sup> But this view seems far less likely to be true than any of the other non-reductive versions of physicalism, like the token-identity theory and functionalism for instance. There are powerful objections against it coming from the so-called multiple realisability of mental (or functional) states and the problem of absent qualia. There is no reason to suppose that kinds of mental events are realised by kinds of physical events such that there could be one-to-one correlations between them. 'Pains', for instance, could be realised in this physical way and then that even though these states themselves have nothing in common apart from simply being physical states. And it is also conceivable that someone could be in exactly the same physical state as I am when feeling pain and yet that person have a completely different mental state. The problem is that the mental properties of events just don't seem to be the sorts of things which could have a physical essence which could be captured by science.<sup>25</sup>

24. Smart 1959, 1963. The other alternative, of course, is simply to deny that there are distinctively mental events and properties at all, the eliminative materialist line. On this view there would be no problem of compatibility at all, for the mentalistic language could simply be done away with. It could at most be regarded instrumentally, as Dennett has argued it should. But this, I think, is wildly implausible.
25. See McGinn 1982:21-36 for more on this.

The view that the mental just is physical is a far harder claim to assess because it just does not seem that we at present can understand what it exactly involves. It is not just that we do not have any clear idea of how my feeling a pain right now can be identical with certain chemical or biological goings-on in a certain region of space-time, but it is also that we do not have any idea of how such apparently objective goings-on can ever amount to what seems to be an essentially subjective phenomena. I am referring, of course, to Nagel's notion of what it is like to be a particular creature or what it is like to experience a pain.<sup>27</sup> It is these features of our world that seem doomed to forever elude capture in any physicalist net that sets out to give an objective or extensional description of what our world is like.

When somebody intentionally performs an action, then common-sense tells us that this behaviour is preceded by or somehow involves a conscious episode in the psychological life of the person so behaving. But it is hard, perhaps impossible, to imagine how such a conscious episode can occur without it involving this subjective feature, this feature of what it is like for the person to intend, desire, believe or act as he does. And this feature of consciousness just does not seem capable of objective or physicalist description. It would seem then that if a piece of behaviour has been correctly identified as an action, then no explanation of that

26. Skillen 1984:526 and also Melchert 1986:274.

27. Nagel 1974. See also H. Robinson 1982:4 where he argues that this is a fatal objection to physicalism of any sort.

behaviour which is consistent with the absence of this feature can be adequate. For this is just part of what it is for a person to act intentionally. It could not be action without this feature.

It is curious that in a recent article defending Davidson's Anomalous Monism from the charge of epiphenomenalism, Norman Melchert should write as follows:

There is no necessity attached to interpreting behaviour under intentional concepts. We may (in principle, at least) view the actions of others (and of ourselves) as simply physical events in a physical world. But, one could say, from the moral point of view one ought to accord others this dignity, to regard them as "intentional systems" with a degree of psychological autonomy (1986:273. He specifically refers to Dennett's notion of an "intentional system").

This, I think is clear enough, is simply another variety of  
28  
Instrumentalism.

For, given what has been argued above, Melchert seems to be saying that we can (if we choose) regard the motions of others as actions, but that regarding them as such is quite consistent with us also regarding them as mere physical events which do not involve any conscious or subjective features. But from what was argued in the previous chapter (on Instrumentalism) and above, I think it is quite clear that no sequence of events which does not involve consciousness can rightly be called intentional action. So while it is maintained that mental properties cannot be nomologically tied to physical properties even though we have a criterion of event identity in terms of which mental

28. Gregory Sheridan (1982:436) counts Davidson among the instrumentalists. This, I think, shows why.

events can be identical with physical events, it would seem that any mechanistic account of behaviour will be compatible with the absence of consciousness (and mentality in general) as a causal factor in the genesis of that behaviour. But such behaviour, I have argued, cannot really be action. And this, I have also argued, is just what the explanatory incompatibilist insists on.

The only way to avoid this incompatibilist conclusion would be to argue that there are after all nomological connections between neurological states and properties and mental states and properties. From our knowledge of these nomological connections we should be able to infer from descriptions of neurological states that an intention of such and such a sort or a pain of such and such a phenomenological character is occurring in the mind of the person whose brain it is. Whether this will ultimately be shown to be possible neither I nor anyone else, I think, has any way of knowing. Davidson<sup>29</sup> has explicitly argued against this possibility. Others have<sup>30</sup> felt less sure. One point of view which many writers on the topic have in common is that in order to even see how this could be possible we would need a vastly enriched conceptual apparatus in order to grasp what precisely this identity comes to. And this would seem to involve our developing a physical understanding of what thoughts, intentions, pains and emotions are, something which we at present just do not<sup>31</sup> have. Charles Taylor argues for a "convergence hypothesis"

29. See his "The Material Mind" (1980E) for instance.

See also Swinburne 1986:192.

30. See Meehl 1966 and Skillen 1984.

31. Nagel 1974; H. Robinson 1982; Melchert 1986.

whereby "the present principle of neurophysiology, and a fortiori those of physics and chemistry would be supplemented by concepts of quite a different kind, in which, for instance, relations of meaning might become relevant to neurological process" (1985A:185).

Skillen (1984:526) suggests that "it seems that the 'functionalist' notions of 'storage' and 'information' may be the models for 'bridging' concepts that enable us to see mental states as physical".<sup>32</sup> But he also warns that "premature attempts to provide a 'rough' account of mind in physical terms will tend to combine vulgar reductionism with backdoor animism as homunculi are released into the machinery in the attempt to give it the kiss of real life".

My own view is a lot more pessimistic (from the would-be compatibilist's point of view). I think that some or other kind of dualism is inevitable, whether this be of mental and physical substances (Cartesian dualism) or mental and physical properties (the token-identity theory) or functional and physical states (functionalism). The case is far harder to prove with intentional or propositional states than it is with those mental items which involve sensations, such as pains, aches, emotions, etc. The reason, as has often been noted,<sup>33</sup> is that the propositional states (thoughts, desires, beliefs, etc.) are less obviously connected with a first-person or subjective perspective than are those mental items which involve the more sensuous side of consciousness. It is

32. Though I have argued against it, this nevertheless seems to be something both exciting and hopeful in Dennett's writing on the topic.

33. For instance by McGinn 1982:19. See also Baker 1980.

far easier, for instance, to think of a computer as 'believing' or 'desiring' than it is to think of it experiencing pain or undergoing some profound emotion. Thus it is far easier for us to think of the terms and concepts of Artificial Intelligence as providing 'bridging concepts' which may eventually enable us to link up processes such as remembering, calculating and even believing with complex physical processes of the sort going on in a computer when we apply these terms to them. But pains and emotions - those states which have distinctive phenomenological features - seem to be a different matter entirely. They seem to involve something irreducibly subjective, and it is hard to see how this feature can be referred to or described in a purely objective language, as physics aims to be. While this remains the case, I would urge, we have no way of seeing how complete mechanism can be true, and so we have reason to believe that complete mechanistic explanations of behaviour are indeed incompatible with that behaviour being genuine action, involving, as it must, the causal efficacy of these conscious episodes.

Something has to give, and since it cannot be our ordinary way of regarding and explaining certain items of behaviour as action,<sup>34</sup> it will have to be the belief that there in fact can be complete mechanistic explanations of all our behaviour. And the only way that this can be secured would be if physical determinism in fact is false, at least insofar as it applies to those of our motions which we regard as our free and rational intentional actions.

33. See Charles Taylor 1985A:174; Norman Malcolm 1968; Flew 1975:116-122.

## CHAPTER FIVE

### CONCLUSION

#### 5.1 Summary of the Argument

In the previous chapters I have presented arguments against the view that intentional explanations of behaviour are compatible with there also being complete mechanistic explanations of that behaviour. In the first chapter it was conceded that there has to be some sort of compatibility between intentional and mechanistic explanations for the simple reason that my intentionally raising my arm most certainly does involve certain physiological occurrences which can be explained mechanistically. The question was whether these explanations, in the case of an intentional action, could be complete, i.e. whether there could be an unbroken chain of physical causes, each being a full explanation of its effect, going indefinitely back to something (an event) outside or independent of the agent whose behaviour it is. This possibility was denied, and the next three chapters were all attempts to defend this conclusion against various sorts of objections.

In Chapter Two I considered the view that intentional and mechanistic explanations are too different to ever be thought of as rivals, and that the reality of human action therefore leaves untouched the possibility of unlimited success in the mechanistic explanation of human motions. This view was rejected for the following basic reasons: firstly, it fails to come to terms with what we all intuitively believe to be

true, namely that our desires, beliefs, reasons and intentions do play a role in bringing about those of our motions which are our actions. Secondly, denying that these items play a causal role led the double-language theorist to a severe problem in accounting for the difference between mere bodily movements and those movements which also are actions. It was argued that the only way these could be distinguished, even at the physical level, was if there is some sort of causal difference in the way these two sorts of bodily movements are in fact generated. Another problem with the double-language theory is that actions often are bodily movements, and if these can be completely explained mechanistically, then so too can the actions. By denying the causal role of intentions and the like the double-language theorist must give up on finding any significant difference between the two sorts of movements, and this seems just plain wrong. Furthermore, it was found that the arguments for denying that intentional explanation is causal explanation cannot be sustained and that there are very good reasons for preferring some or other version of the causal theory. It enables us to significantly distinguish between actions and mere movements which the double-language theory cannot do.

In the next chapter I discussed the Instrumentalist version of compatibilism, but it was argued that this is entirely unsatisfactory because it fails to give an adequate account of the sort of rationality involved in intentional action. Intentional explanations, it was concluded, come in two varieties, namely those that do and those that do not make assumptions as to whether the system whose behaviour is being



explained is in fact a conscious agent. Instrumentalism establishes the compatibility with mechanism of only the latter sort of intentional explanation, but it is rather the former sort of intentional explanation, that which ascribes genuine conscious states, that is involved in our ordinary understanding and explanation of actions. It was concluded that the only viable form of compatibilism would be that which attempts to find systematic connections between conscious states and physical states.

In the chapter on Physicalism this last suggestion was considered, and the negative conclusion to emerge from that discussion was that both the Correlation-thesis and Token-Identity versions of physicalism fail to make the mental as such causally efficacious and hence reduce to a type of epiphenomenalism. And epiphenomenalism, I take it, though possibly true, is extremely counter-intuitive. It in any case amounts to the view that there are no actions (as they are traditionally conceived), but only physical events which may or may not be accompanied by conscious occurrences or mentality. The only remaining possibility, I concluded, was that consciousness and the mental as such (with all its subjective features and phenomenological properties) are simply physical goings-on which nevertheless are mental as well. But I suggested that this view, though probably false, is in any case far from being understood. While it may, I conceded, ultimately be vindicated by conceptual advances to be made in the future,<sup>1</sup> we nevertheless may be justified now in believing that incompatibilism is true.

1. Popper (1983:205) calls this "promissary materialism".

## 5.2 Actions, Agents and Other Minds

My aim throughout this discussion has not been to put forward a positive thesis about human action and free will, but has been concerned with the more modest aim of defending explanatory incompatibilism, what I take to be the common-sense view, against certain powerful and sophisticated objections. That task has now more or less been completed. The conclusion we have reached is that intentional explanations are indeed incompatible with complete mechanistic explanations, and the basic reason for this is that complete mechanism, which entails physical determinism, fails to find room in its austere picture for what we all intuitively accept to be true, namely that our mental lives as such make a causal difference to the way the physical world is. In order to make room for intentional action, we would need to give up the view that there in fact can be complete mechanistic explanations of all our behaviour. And the only way that this can be possible is if physical determinism in fact is false, as libertarians have always insisted that it is. That this effectively blocks complete mechanism is evident from the fact that if physical indeterminism is true, then there cannot be full or complete mechanistic explanations for everything that happens in the physical world, but only partial explanations which can predict and explain to within only a limited degree of accuracy.<sup>2</sup>

Once this has been admitted, then a picture of how human

2. This was argued for and established in Chapter One. But see also Thorp 1980:100.

agency and action is possible begins to emerge, and it is a picture which neither denies the reality of human free action nor places intolerable restrictions on what we may want to achieve insofar as a detailed mechanistic account of human behaviour is concerned.

And while I do not want to attempt a rigorous and systematic argument or proof that this indeed is the correct picture, it nevertheless is incumbent on me to say at least something on what I take this picture to be and how it in fact is possible. To this I now turn.

The question that needs to be answered is what the necessary and sufficient conditions are for a piece of behaviour being an intentional action. So let us construct an imaginary case and pursue the question step by step. Let us suppose that we observe some behaviour which we then characterise and explain as John's raising his arm in order to signal the waiter.

What would make that behaviour an intentional action ?

The first condition to be met, one about which most writers would agree, must at least be that there is an intentional explanation of the behaviour. This in a way is obvious. We normally count behaviour as action rather than as mere movement not by examining how the behaviour was physically produced, but by seeing whether it makes sense in the context to speak of reasons the agent might have had in behaving as he did. And in the example above this condition is obviously satisfied - we say that John raised his arm because he wanted to signal the waiter, and this is an intentional explanation.

But having said this a complication immediately arises. For

in the discussion of Dennett's instrumentalist version of compatibilism it was conceded that there is a kind of intentional stance we can adopt toward a system which does not commit us to regarding that system as a conscious agent or as really having a mental life at all. This is what we do when we explain the behaviour of a computer by saying that it wants this or desires that. But it has also been argued throughout this discussion that having a mental life is absolutely necessary for being an agent or for performing intentional actions. And this implies that there merely being an intentional explanation of behaviour is not enough to confer on a piece of behaviour the status of action. What, for example, if in our imaginary case John in fact is a zombie with no mental life at all? Even though we can explain and understand his motions as though it was an action he performed with the intention so to do, this does not establish his behaviour as action. Something more is needed, something which guarantees, as it were, that John is a conscious agent.

Now it has been argued that this 'something' comes in the form of Davidson's insistence that intentional explanations not only rationalise the actions which they explain, but also that the reasons cited be the reasons on which the agent acted.<sup>3</sup> When we say that John raised his arm because he wanted to signal the waiter and he believed that by raising his arm he would do so, then the desire and belief (Davidson's 'primary reason') cited must not only justify the behaviour, but they must also be the causes of the behaviour.

3. Davidson 1980A, 1980C.

In this way it is supposed one can get mental events into the picture. If it really is true that John's desire and his belief caused his behaviour, then this will ensure not only that John is a conscious agent, but also that his being conscious somehow contributed to his behaviour. He wanted to do something, and this wanting caused him to do it. And this would make his behaviour an intentional action.

Now no picture like this can be adequate. And there are two basic reasons for this. The first, which we have already encountered in the discussion of Davidson's view, is that this picture leads straight to the epiphenomenalism of the mental, at least that is if one supposes as Davidson does that the mental events which cause the physical motions are identical with physical events in the brain. This was argued extensively in the previous chapter. But even if one supposed that the mental events were somehow non-physical (causing the appropriate physical events just at those crucial points where they are not wholly caused by antecedent brain states), even then this picture could not work. And the reason is, as Davidson himself points out, that a person's desires and beliefs may cause his behaviour but in a totally weird or deviant way such that we would not call the behaviour an intentional action. This is the so-called Problem of Deviant Causal Chains. Davidson offers his own  
4  
example:

A climber may want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief

4. For similar examples see McGinn 1979:33 and Richard Taylor 1966:249.

and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally (1980J:79).

In this case the desires and beliefs rationalise the behaviour and they also cause it, but they cause it in an uncharacteristic or deviant manner such that the resulting behaviour is not an intentional action. But how, we can fairly ask, is the desire-belief version of the causal theory to exclude cases such as these? How is it to give an account of what the 'characteristic' or 'right' way is in which desires and beliefs are meant to cause actions?

Goldman has written:

To this question, I confess, I do not have a fully detailed answer. But neither do I think that it is incumbent on me, qua philosopher, to give an answer to this question. A complete explanation of how wants and beliefs lead to intentional action would require extensive neurophysiological information, and I do not think it is fair to demand of a philosophical analysis that it provide this information (1970:62).

But one has the distinct feeling that Goldman is here simply passing the buck. And this, as Swinburne and Bishop have also pointed out, is entirely unsatisfactory. Bishop writes:

Now, it is obvious that our concept of intentional action has some empirical features. And it could hardly be denied that there are important neuro-physiological differences between genuine intentional action and its deviant semblance which are still to be discovered. But the event-causalist [(by which Bishop means the desire-belief theorist)] is here envisaging that there are features in our concept of non-deviant intentional action which are both empirical and unknown. But this cannot be. If one makes some distinction on empirical grounds, then we must know something at least of what those empirical grounds are (1983:68).

Bishop, I think, has a point here. The difference between intentional action and its deviant semblance is something

that concerns our explanations of behaviour at the ordinary and everyday level of intentional explanation, for it is at this level that we understand or grasp that there is a difference here. It is not as though the difference is a yet to be verified empirical one. Swinburne makes more or less the same point in his comment on the passage from Goldman above:

But this really will not do. For hundreds of years men have been able to distinguish, among cases where wants cause the events wanted, those cases where an action was performed. We have distinct concepts which we know how to apply. It is indeed incumbent upon a philosopher to analyse the difference - although it is not up to him to say which neurophysiological goings-on are physically necessary to produce a case of an action being performed. Goldman has not analysed the difference (1979:41n).

I would conclude that although there is something very right about this version of the causal theory, the belief-desire model nevertheless fails to offer an account of the sufficient conditions of intentional action. In our own story constructed before this digression, it might well be the case that John's desire to signal the waiter is so great that it makes him so anxious with the result that a muscular spasm causes his arm to go up. And this seems to indicate that it cannot simply be a matter of the desires and beliefs causing the actions they rationalise - they must also cause it in the right way. But the desire-belief theorist, I think, will never come up with a philosophical  
5  
account of what that right way is. The problem, I would

5. Christopher Peacocke (1979) has apparently developed an account of intentional action which is said to hold out considerable hope of providing just the sort of deviance-excluding analysis which is required for the belief-desire theory. I must quite frankly confess that I do not understand his argument. But see Bishop 1981 and 1983 for a discussion and rejection of his theory.

suggest, is that no clauses can be added to the basic view which do not, in some way or another, make a reference to the agent's causing something or other that is required for intentional action.<sup>6</sup> But I will say something more on this in a short while.

At the moment I want to proceed with our search for the necessary and sufficient conditions of intentional action, and the position we have reached is roughly as follows: firstly, there must be no complete mechanistic account of the behaviour, because this will exclude there also being causally efficacious mental antecedents of the behaviour, and this, I have argued throughout, is absolutely essential for intentional action. The next condition that has to be met is that there must be an intentional explanation of the behaviour. But this condition needs to be bolstered in order to exclude those cases where there may be an intentional explanation of the behaviour which is quite compatible with their being no mentality involved at all. The desire-belief account of intentional explanation was thought to provide this extra condition, but it is also not adequate for one of two reasons or both: either it leads to the inefficacy and redundancy of the mental and it faces insuperable problems because of the possibility of deviant causal chains, or, in the case of mental events not being identical with physical events, it still faces problems connected with deviant causal chains. Something more is definitely needed, something which does not have problems of deviant causal chains and which also ensures the efficacy of the mental.

6. For a detailed account of this see Bishop 1983.



Now this something, I believe, comes in the form of the theory of agency. It says that those motions which are intentional actions are caused, not by antecedent mental events or physical events, but by the agents whose actions they are. This satisfies all our conditions to date. It satisfies the condition that there be no complete mechanistic explanation of the behaviour, because there would be no need for introducing the concept of agent-causation if all our motions were fully caused by antecedent physical events. It also satisfies the condition that there be an intentional explanation of the behaviour. We say that John raised his arm because he wanted to signal the waiter, and by this we could understand that John brought about or caused the relevant motions of his body which constituted his signalling, and he brought this about because he wanted X and believed that by Y-ing he would get X. John's reasons here can be thought of as causes which do not operate on his body or his brain, but on him.

What is right about the belief-desire theory is that it attempts to get the mental into the causal picture, but it fails to do this in a satisfactory way. For not only is there the problem of wayward causal chains, but it appears unable to do justice to the rationality which is involved in intentional action. What I mean is this: for a piece of behaviour to be an intentional action, on this belief-desire view, the agent's belief and desire must not only rationalise and cause the behaviour, but it also has to be true that these beliefs and desires rationalise the behaviour from the agent's point of view. This was argued in the discussion of Instrumentalism where I drew the distinction between designed and enacted-rationality. Take our case above. Suppose that

John wants to signal the waiter and he also believes that by raising his arm he would signal. Suppose that John also raises his arm. Now if the desire and belief are to cause the behaviour such that it is an intentional action, then they cannot do so as disjointed or separate events which are held, as it were, in separate compartments of his mind. In order for him to be rational in behaving as he does these two intentional states somehow have to come together in his consciousness. He must see the logical connection between them. But it is this fact about what it is to be rational in action which, I think, stands opposed to any attempt to analyse action in purely event-causal terms. The 'right' way in which desires and beliefs must cause the actions they rationalise must be through or via the consciousness of the agent who acts. It is here, in the consciousness of the agent, that the rationality is to be found. It is this, I believe, which lies behind Richard Taylor's view when he writes:

When I believe that I have done something, I do believe that it was I who caused it to be done, I who made something happen, and not merely something within me, such as one of my own subjective states, which is not identical with myself. If I believe that something not identical with myself was the cause of my behaviour - some event wholly external to myself, for instance, or even one internal to myself, such as a nerve impulse, volition, or what-not - then I cannot regard that behaviour as an act of mine, unless I further believe that I was the cause of that external or internal event (1983:48).

The claim I am making is that no number of clauses excluding deviance in the causal chain from intentions to behaviour can work if they do not at some point refer to the agent's doing something - his believing, his desiring, his noticing, etc. How else can it be the action of an agent? Surely not just by being caused by events which somehow involve the agent!

What is required is the agent's own involvement, and I see no way of breaking this up into separate events which can then somehow be stuck together so as to make an action. It is as absurd, I believe, as trying to analyse the notion of 'activity' by saying that it is essentially constituted by a group of separable events which are all intrinsically passive.<sup>7</sup>

I take it then that we have to introduce some such notion as that of agent-causality.<sup>8</sup> What we have then is the view that a piece of behaviour is an intentional action if it was brought about by an agent meaning so to do. The 'meaning so to do' can be understood as saying that the agent did what he did because he wanted to do it, where this can be understood as saying that the want caused him to do it. And it is clear that no problem of deviant causal chains can arise in such a case. For if the agent did it, even in the case of non-basic action where an agent does something by doing something else, then it would be true that no other event caused whatever it was that came about. In Davidson's example of a deviant causal chain involving the climber who wants to rid himself of the danger of holding another man on a rope, we may ask what exactly it is that the agent did. He desired and he believed to be sure, but what action did he perform? The

7. See Thorp 1980:110-115 for a fascinating discussion of this issue.

8. This idea has been advocated in recent years most notably by Chisholm (1976;1982) and Richard Taylor (1966;1983). Others who have defended something like it include Wiggins (1973), Popper and Eccles (1983), Thorp (1980), Bishop (1983), Ferre (1973) and then even Aristotle, Aquinas and Thomas Reid. See Chisholm (1976) for these historical claims.

answer, of course, is no action at all. And the reason is clear. There was no instance of agent-causality in this case.

The virtues of introducing a concept such as agent-causation are clear. It helps the libertarian find a middle-ground between simple indeterminism or randomness and causal determination, for instead of actions being causally determined by antecedent events or being simply uncaused as such, the agent-causal account of action says that those of our bodily motions which are actions are indeed caused, but caused by the agents whose actions they are. This then constitutes a reply to Ayer's view that the libertarian is caught in an inescapable dilemma: either actions are causally necessitated by antecedent events or they are the result of random and capricious events, and in neither case can they be viewed as free and responsible actions.

But the agent-causalist can reply that of course the alternatives that Ayer suggests are not exhaustive. Though physically non-necessitated, our actions can indeed be caused by us in such a way that they are intelligible at the intentional rather than the physical level.

But the disadvantages of the view are also quite clear. For it involves us postulating a rather obscure notion of, causality alongside the event-causal notion which we seem to know and handle so well. And if anything, the notion of agent-causation is indeed mysterious. How is it supposed to

9. Ayer 1982:10.

work ? And how are we supposed to detect its presence ? Why should we just not say that an event is simply uncaused rather than say that it is physically uncaused but agent-caused all the same ?

Before attempting something like an answer to these questions, a rather long digression reviewing once again some of the ground we have covered will be in order.

We have already seen that physical indeterminism is a necessary condition for intentional action. And we have also seen that intentional intelligibility is another. Might these two conditions not be enough for something being an intentional action ?

David Wiggins has suggested that perhaps all we need to regard behaviour as the free and intentional action of an agent is that it come about non-deterministically but that it nevertheless be "coherent and intelligible in the low-level terms of practical deliberation", i.e. that it submit to successful intentional explanation.<sup>10</sup>

Now while Wiggins is suggesting two conditions which, I believe, are necessary for intentional action, these two conditions by themselves cannot be jointly sufficient for something being an intentional action. And the reason is that we may suppose it to be possible that there be a physically indeterministic system which even though it is utterly devoid of mentality or consciousness, can nevertheless be described and (partly) explained by adopting the intentional stance toward it.

10. Wiggins 1973:52.

Being entirely devoid of mentality, it cannot rightly be said to act at all. But how are we meant to know this? How are we meant to know that such a system which can be intentionally described and explained is nevertheless devoid of mentality? Wouldn't there being an intentional explanation of its behaviour together with the fact that it is physically indeterministic constitute sufficient evidence for us to believe that it 'has a mind' and is therefore capable of genuine action?

The answer to this, I believe, should be in the affirmative. For it would be too strange, too much of a coincidence if a physically indeterministic system also behaved in such a way that it could be understood and interacted with by adopting the intentional stance toward it if it was not also the case that the system operated from within the perspective of its own individual psychology. Something like this, I believe, is probably true. It would be too much of a brute coincidence if a system was not 'minded' but was also physically indeterministic and intentionally intelligible. But we are not to confuse evidencing conditions with criteria of identity.<sup>11</sup> It may be the case that the only evidence we have for an object's possession of X is its possession of Y, but we are not to be tempted into saying that its possession of X is therefore constituted by its possession of Y, that is unless we knew that an object could not possibly possess X if it did not also possess Y. But in the case we are discussing this does not seem to be the case. That a system can be described and understood by adopting the intentional stance

11. Swinburne 1986:155.

toward it is evidence that it 'has a mind', but its having a mind is not constituted by this fact. We can readily imagine cases - computers, for instance - where this is not sufficient for a system in fact 'having' or 'being a mind'. That a system is also physically indeterministic would be further evidence that it is 'minded', but its being 'minded' is also not constituted by these two conditions.

The importance of these considerations comes out when it is observed that some contemporary eliminative materialists would have us believe that 'having a mind' simply consists in having one's behaviour successfully predicted and explained in terms of intentional states attributed to one. Paul Churchland sums up the view as follows:

[T]he hypothesis that a specific individual has conscious intelligence is also an explanatory hypothesis, on this view. And it is plausible to the degree that the individual's continuing behaviour is best explained and predicted in terms of desires, beliefs, perceptions, emotions, and so on. Since this is, in fact, the best way to understand the behaviour of most humans, one is therefore justified in believing that they are 'other minds'. And one will be similarly justified in ascribing psychological states to any other creatures or machines, so long as such ascriptions sustain the most successful explanations and predictions of their continuing behaviour (1986:71).

But what is meant here by "most successful explanations" and "best way to understand"?

I argued in Chapter One that the best explanation of an event is the one that renders that event most strongly in explanation, and it seems clear that mechanistic explanations of phenomena, insofar as they are based on deterministic laws, would be precisely those that render an event most strongly in explanation. So where these are available, then any other "mentalistic" or intentional explanations could not

possibly be the best and most successful explanations we have. And this, I think, implies that one cannot really believe that a system is 'minded' if one believed that there in fact are as yet undiscovered deterministic laws which could feature in a complete mechanistic account of its behaviour. For believing that there is such a complete mechanistic explanation of a system's behaviour would be enough reason to doubt that it has any mentality at all. Of course none of this will give any discomfort to the eliminative materialist, for he believes that the hypothesis of 'other minds' can be superceded by such a comprehensive mechanistic theory. But it should give us pause for reflection. For the only way that we can continue to believe in 'other minds' in the same sense that we do now, would be for us to fail to find such a complete mechanistic account of human behaviour. The reason is that it is precisely physical indeterminism which allows for the possibility of our intentional explanations being hegemonic with respect to any underlying mechanistic explanations.

In reply to the various questions posed about agent-causation a few paragraphs back, the rudiments of an answer begin to emerge.

Firstly, we can say that we need to postulate something like agent-causality in order to adequately account for the data. Decomposing action into a series of passive and separate events will not work.

Secondly, it cannot be expected that we give an analysis of agent-causation if by this is understood saying what the various elements are of which it is composed.

This is precisely what we cannot do, for it would be trying



to understand agent-causality on the model of event-causality. All we can say is that sometimes an agent brings about physical events, presumably events in his brain, which events are then the cause of the various motions which are his intentional action. Thus when John raises his arm in order to signal the waiter, and if this indeed is an intentional action, then John intentionally brings it about that his arm goes up by bringing it about that certain events in his brain cause the motion of his arm. He does not intentionally bring about the events in his brain - he might not even know that he has a brain - but if it is true that he raised his arm, then it must be true that he and not some other event, even an event involving him such as a desire or a nervous impulse, is the cause of those physical events which resulted in his arm's going up.

All of this, of course, is still very obscure. But that cannot be an objection to it in itself. Chisholm has argued that the very same obscurity and mystery surrounds the concept of one event causing another as well. He says:

The analogous problem, which holds for 'transeunt causation', or causation by an event, is this: 'What is the difference between saying, of two events A and B, that B happened and then A happened, and saying that B's happening was the cause of A's happening?' And the only answer that one can give is this - that in the one case the agent was the cause of A's happening and in the other case event B was the cause of A's happening. The nature of transeunt causation is no more clear than is that of immanent [(by which he means agent)] causation (1982:31).

Chisholm, I think, has a point here. How indeed are we to understand the notion of a 'cause', any sort of 'cause', at all? All we can say is that one event somehow produces another event, and we have our stock of words which describe this transaction - pushings, pullings, forcings, makings

happen and so forth - but none of them, it seems, can specify the essence of this mysterious notion without in some way or another involving other notions which already contain the primitive notion of a 'cause'.<sup>13</sup>

Let us rest with the idea that it is unanalysable.

But how are we to detect the presence of a case of agent-causality as opposed to a simple case of uncausality? The answer to this, if all that has gone before in this discussion is coherent, is clear. For one good reason for saying that a piece of behaviour is caused by an agent rather than being uncaused is just that there is an intentional explanation of its behaviour and its behaviour is physically undetermined. It would be too much of a coincidence, I have suggested, if a system's behaviour was in fact physically undetermined but nevertheless capable of being understood in the normal language of human action and its explanation. So a case of genuine intentional action - movement brought about by an agent - can be distinguished from a case of mere movement by finding (a) that the behaviour cannot be completely explained mechanistically, and (b) by there being an intentional explanation of the behaviour which is hegemonic with respect to any partial mechanistic explanation of that behaviour. And this is why the behaviour of a quantum particle, or of some or other indeterministic machine, can never be an intentional action - because an

12. See Holland 1980:211-228.

13. Anscombe discusses this in her Inaugural Lecture at Cambridge University. See Anscombe 1981B.

14. See also Richard Taylor 1966:91-92, 112 and Thorp 1980:106.

intentional explanation of its behaviour would leave us as much in the dark as before and it is safer, more explanatory, to stick to the physical level of explanation in these cases. And this also explains why there would be no reason to regard computers or any other machines we may create as agents or 'other minds'. The reason is not that we cannot adopt the intentional stance toward them, for this we can do, but rather that we will know that their behaviour can be completely explained mechanistically - for how else are we to construct them if not in accordance with physical laws? - and this would be enough reason to doubt, so I have argued throughout, that they had any mentality or consciousness at all.

The picture I am suggesting obviously has a dualistic ring to it. But it need not be seen as a twentieth century version of Cartesian dualism which seeks crucial gaps of randomness in the causal framework of the world, in our brains, through which causally efficacious immaterial agents could be sneaked into the picture.

Agents, on the view I am suggesting, are simply the flesh and blood persons we know ourselves to be. Being creatures of flesh and blood, we form part of the causal fabric of the world which it is the business of physics, chemistry, biology and even ecology to describe and explain. But it would simply be dogmatism to insist that all behaviour that ever was or ever will be can be adequately explained from within these levels of description alone. If physical indeterminism is true, as I have assumed it is, then that view crumbles. The full picture of what constitutes the human reality cannot be described from these levels alone. And if these levels

are not sufficient to capture the full picture, does ascending to higher or different levels therefore commit us to postulating a separate reality of immaterial minds or agents which from time to time interfere in this physical world? I think not. All we need to realise is that we have a multi-levelled reality which requires a multi-levelled ontology in order to adequately explain and describe its richness. We need to see, as Charles Taylor says, "that although some principles govern the behaviour of all things, others apply only to some; and yet the latter cannot be shown as special cases of the former" (1985A:186).

Agents are creatures of flesh and blood, but the fullness of their agency will be missed if looked at from only the physical point of view. The pattern emerges at a different level, namely the level of our ordinary and common-sense view of persons. But this pattern can only emerge, so I have tried to argue, if every element of the total picture is not determined by antecedent physical causes which go back indefinitely. At some point a different principle is in operation, the principle of agency, and failure to see it will result in a failure to see the relevant pattern.

---

## BIBLIOGRAPHY

Note: The date after the author's name indicates the date of the publication used and not that of the original publication.

- (1) ANSCOMBE, G.E.M. 1979. Intention. Basil Blackwell, Oxford.
- (2) ANSCOMBE, G.E.M. 1981. Collected Philosophical Papers 2: Metaphysics and the Philosophy of Mind. Basil Blackwell, Oxford.
- (3) ANSCOMBE, G.E.M. 1981A. "A Reply to C.S. Lewis' Argument that Naturalism is Self-Defeating" in (2) above.
- (4) ANSCOMBE, G.E.M. 1981B. "Causality and Determination" in (2) above.
- (5) ARMSTRONG, D.M. 1981. The Nature of Mind. Harvester Press, Sussex, England.
- (6) AYER, A.J. 1964. The Concept of a Person and Other Essays. MacMillan and Co, London.
- (7) AYER, A.J. 1982. "Freedom and Necessity" in (114) below.
- (8) BAKER, L.R. 1981. "Why Computers Can't Act" in American Philosophical Quarterly, 18:2.
- (9) BENNETT, J. 1976. Linguistic Behaviour. Cambridge University Press, Cambridge.
- (10) BISHOP, J. 1981. "Peacocke on Intentional Action" in Analysis, 41.
- (11) BISHOP, J. 1983. "Agent-Causation" in Mind 92.
- (12) BODEN, M.A. 1978. Purposive Explanation in Psychology. Harvester Press, Sussex.

- (13) BOYLE, J. 1976. (with GRIZES, G. and TOLLEFSEN, O.). Free Choice: A Self-Referential Argument.  
University of Notre Dame Press, Notre Dame, Indiana.
- (14) CHISHOLM, R. 1976. "The Agent as Cause" in Action Theory.  
(eds. Brand, M. and Yolton, D.) Reidel, Dordrecht.
- (15) CHISHOLM, R. 1982. "Human Freedom and the Self" in  
(114) below.
- (16) CHURCHLAND, P. 1986. Neurophilosophy: Toward a Unified Science of the Mind-Brain. Bradford Books, Cambridge, Massachusetts.
- (17) CHURCHLAND, P. 1986. Matter and Consciousness.  
M.I.T. Press, Cambridge, Massachusetts.
- (18) DANTO, A. 1965. "Basic Actions" in American Philosophical Quarterly, 2.
- (19) DAVIDSON, D. 1980. (1) Essays on Actions and Events.  
Oxford University Press, Oxford.
- (20) DAVIDSON, D. 1980A. "Actions, Reasons and Causes" in  
(19) above.
- (21) DAVIDSON, D. 1980B. "Mental Events" in (19) above.
- (22) DAVIDSON, D. 1980C. "Agency" in (19) above.
- (23) DAVIDSON, D. 1980D. "Intending" in (19) above.
- (24) DAVIDSON, D. 1980E. "The Material Mind" in (19) above.
- (25) DAVIDSON, D. 1980F. "Psychology as Philosophy" in (19)  
above.
- (26) DAVIDSON, D. 1980G. "Hempel on Explaining Action" in  
(19) above.
- (27) DAVIDSON, D. 1980H. "Causal Relations" in (19) above.
- (28) DAVIDSON, D. 1980I. "The Individuation of Events" in  
(19) above.

- (29) DAVIDSON, D. 1980J. "Freedom to Act" in (19) above.
- (30) DAVIDSON, D. 1982. "Paradoxes of Irrationality" in Philosophical Essays on Freud. (eds. Wollheim, R. and Hopkins, J.) Cambridge University Press, Cambridge.
- (31) DAVIDSON, D. 1985. "Deception and Division" in Actions and Events: Perspectives on the Philosophy of Donald Davidson. (eds. LePore, E. and McClaughlin, B.) Basil Blackwell, Oxford.
- (32) DENNETT, D. 1973. "Mechanism and Responsibility" in (49) below.
- (33) DENNETT, D. 1979. Brainstorms. Hassocks, Sussex.
- (34) DENNETT, D. 1986. "I Could Not Have Done Otherwise - So What?" in Journal of Philosophy 81.
- (35) DRETSKE, F. 1985. "Machines and the Mental" in Proceedings and Addresses of the American Philosophical Association 59:1.
- (36) ECCLES, J. 1983. (with POPPER, K.) The Self and its Brain. Routledge and Kegan Paul, London.
- (37) ECCLES, J. 1984. The Human Mystery. Routledge and Kegan Paul, London.
- (38) EWIN, R.E. 1968. "Actions, Brain-Processes and Determinism" in Mind 77.
- (39) FERRE, F. 1973. "Self-Determinism" in American Philosophical Quarterly 10:3.
- (40) FLEW, A. 1978. A Rational Animal. Oxford University Press, Oxford.
- (41) FODOR, J. 1968. Psychological Explanation. Random House, New York.
- (42) FRANKFURT, H. 1978. "The Problem of Action" in American Philosophical Quarterly 15:2.

- (43) GOLDMAN, A. 1970. A Theory of Human Action. Prentice-Hall, New Jersey.
- (44) HAMPSHIRE, S. 1959. Thought and Action. Chatto and Windus, London.
- (45) HEMPEL, C. 1965. Aspects of Scientific Explanation. Free Press, New York.
- (46) HEMPEL, C. 1966. Philosophy of Natural Science. Prentice-Hall, New Jersey.
- (47) HESS, P.H. 1981. "Actions, Reasons and Humean Causes" in Analysis 41.
- (48) HOLLAND, R.F. 1980. Against Empiricism. Basil Blackwell, Oxford.
- (49) HONDERICH, T. (ed.) 1973. Essays on Freedom of Action. Routledge and Kegan Paul, London.
- (50) HONDERICH, T. 1982. "The Argument for Anomalous Monism" in Analysis 42.
- (51) HONDERICH, T. 1983. "Anomalous Monism: Reply to Smith" in Analysis 43.
- (52) HONDERICH, T. 1984. "Smith and the Champion of Mauve" in Analysis 44.
- (53) KENNY, A.J.P. 1963. Action, Emotion and Will. Routledge and Kegan Paul, London.
- (54) KENNY, A.J.P. 1973. "Freedom, Spontaneity and Indifference" in (49) above.
- (55) KIM, J. 1966. "On the Psycho-Physical Identity Theory" in American Philosophical Quarterly 3.
- (56) LEVIN, M. 1979. Metaphysics and the Mind-Body Problem. Oxford University Press, Oxford.
- (57) LEWIS, C.S. 1964. Miracles. Collins Fontana, London.



- (58) LONG, D. 1979. "Agents, Mechanisms and Other Minds" in Body, Mind and Method. (eds. Gustafson, D. and Tapscott, B.) Reidel, Dordrecht.
- (59) LUCAS, J.R. 1970. The Freedom of the Will. Oxford University Press, London.
- (60) MACINTYRE, A. 1957. "Determinism" in Mind 66.
- (61) MACINTYRE, A. 1960. "Commitment and Objectivity" in The Sociological Review no. 3.
- (62) MADELL, G. 1986. "Neurophilosophy: A Principled Sceptic's Response" in Inquiry 29:2.
- (63) MALCOLM, N. 1968. "The Conceivability of Mechanism" in The Philosophical Review 78:1.
- (64) MATSON, W. 1976. Sentience. University of California Press, Berkeley.
- (65) MCGINN, C. 1979. "Action and its Explanation" in Philosophical Problems in Psychology. (ed. Bolton, N.) Methuen, London.
- (66) MCGINN, C. 1982. The Character of Mind. Oxford University Press, Oxford.
- (67) MEEHL, P. 1966. "The Compleat Autocerebroscopist" in Mind, Matter and Method. (eds. Feyerabend, P. and Maxwell, G.) University of Minnesota Press, Minneapolis.
- (68) MELCHERT, N. 1986. "What's Wrong with Anomalous Monism?" in Journal of Philosophy 83:5.
- (69) MELDEN, A.I. 1961. Free Action. Routledge and Kegan Paul, London.
- (70) MEYERSON TAYLOR, D. 1982. "Actions, Reasons and Causal Explanations" in Analysis 42.
- (71) MULLANE, H. 1971. "Psychoanalytic Explanation and Rationality" in Journal of Philosophy 68.

- (72) NAGEL, E. 1971. The Structure of Science. Routledge and Kegan Paul, London.
- (73) NAGEL, T. 1974. "What is it like to be a Bat?" in The Philosophical Review 83.
- (74) NAGEL, T. 1979. "Moral Luck" in Mortal Questions. Cambridge University Press, Cambridge.
- (75) O'CONNOR, D.J. 1971. Free Will. Doubleday Anchor, New York.
- (76) PEACOCKE, C. 1979. Holistic Explanation: Action, Space, Interpretation. Clarendon Press, Oxford.
- (77) PEARS, D. (ed.) 1963. Freedom and the Will. MacMillan, London.
- (78) PEARS, D. 1973. "Rational Explanation of Actions and Psychological Determinism" in (49) above.
- (79) PETERS, R.S. 1958. The Concept of Motivation. Routledge and Kegan Paul, London.
- (80) PLATO 1954. The Last Days of Socrates. Penguin.
- (81) POPPER, K. 1983. (with ECCLES, J.) The Self and its Brain. Routledge and Kegan Paul, London.
- (82) PUTNAM, H. 1974. "Reductionism and the Nature of Psychology" in Cognition 2:1.
- (83) RICHMAN, R.J. 1969. "Reasons and Causes: Some Puzzles" in Australasian Journal of Philosophy 47:1.
- (84) ROBINSON, H. 1982. Matter and Sense. Cambridge University Press, Cambridge.
- (85) ROBINSON, W. 1979. "Do Pains Make a Difference to our Behaviour?" in American Philosophical Quarterly 16:4.
- (86) RORTY, R. 1965. "Mind-Body Identity, Privacy and Categories" in Review of Metaphysics 19.

- (87) RYLE, G. 1980. The Concept of Mind. Penguin, Harmondsworth, England.
- (88) SCARROW, D. 1981. "The Causality of Reasons: A Survey of Some Recent Developments in the Mind-Body Problem" in Metaphilosophy 12:1.
- (89) SHAFFER, J. 1968. Philosophy of Mind. Prentice-Hall, Englewood Cliffs, New Jersey.
- (90) SHERIDAN, G. 1982. "Can There be Moral Subjects in a Physicalistic Universe" in Philosophy and Phenomenological Research 43.
- (91) SHUTTE, A. 1984. "The Refutation of Determinism" in Philosophy 59:230.
- (92) SKILLEN, A. 1984. "Mind and Matter: A Problem that Refuses Dissolution" in Mind 93.
- (93) SMART, J.J.C. 1959. "Sensations and Brain Processes" in The Philosophical Review 68.
- (94) SMART, J.J.C. 1963. Philosophy and Scientific Realism. Routledge and Kegan Paul, London.
- (95) SMITH, P. 1981. "Hess on Reasons and Causes" in Analysis 41.
- (96) SMITH, P. 1984. "Anomalous Monism and Epiphenomenalism: A Reply to Honderich" in Analysis 44.
- (97) STEINER, M. 1986. "Events and Causality" in Journal of Philosophy 83:5.
- (98) STRAWSON, P.F. 1980. Freedom and Resentment and Other Essays. Methuen, London.
- (99) SUPPES, F. 1974. Probabilistic Metaphysics. University of Uppsala Press, Uppsala.
- (100) SWINBURNE, R. 1979. The Existence of God. Oxford University Press, Oxford.

- (101) SWINBURNE, R. 1982. "Are Mental Events Identical with Brain Events" in American Philosophical Quarterly 19.
- (102) SWINBURNE, R. 1986. "The Indeterminism of Human Actions" in MidWest Studies in Philosophy 10.
- (103) SWINBURNE, R. 1986A. The Evolution of the Soul. Clarendon Press, Oxford.
- (104) TAYLOR, C. 1964. The Explanation of Behaviour. Routledge and Kegan Paul, London.
- (105) TAYLOR, C. 1970. "The Explanation of Purposive Behaviour" in Explanation in the Behavioural Sciences. (eds. Borger, R. and Cioffi, F.) Cambridge University Press, Cambridge.
- (106) TAYLOR, C. 1985. Human Agency and Language: Philosophical Papers vol.1. Cambridge University Press, Cambridge.
- (107) TAYLOR, C. 1985A. "How is Mechanism Conceivable" in (106) above.
- (108) TAYLOR, C. 1985B. "Cognitive Psychology" in (106) above.
- (109) TAYLOR, R. 1966. Action and Purpose. Prentice-Hall, Englewood Cliffs, New Jersey.
- (110) TAYLOR, R. 1983. Metaphysics. Prentice-Hall, Englewood Cliffs, New Jersey.
- (111) THORP, J. 1980. Free Will: A Defence Against Neurophysiological Determinism. Routledge and Kegan Paul, London.

- (112) TOULMIN, S. 1970. "Reasons and Causes" in Explanation in the Behavioural Sciences. (eds. Borger, R. and Cioffi, F.) Cambridge University Press, Cambridge.
- (113) VAN INWAGEN, P. 1983. An Essay on Free Will. Oxford University Press, Oxford.
- (114) WATSON, G. (ed.) 1982. Free Will. Oxford University Press, Oxford.
- (115) WIGGINS, D. 1973. "Towards a Reasonable Libertarianism" in (49) above.
- (116) WINCH, P. 1980. The Idea of a Social Science. Routledge and Kegan Paul, London.
- (117) WHITELEY, C.H. 1968. "Mental Causes" in Royal Institute of Philosophy Lectures vol.1 (1966-67) : The Human Agent.
- (118) WHITELEY, C.H. 1973. Mind in Action. Oxford University Press, Oxford.
- (119) WRIGHT, L. 1976. Teleological Explanations. University of California Press, Berkeley.