FREEDOM AND THE WILL

In the same series

*

DAVID HUME: A SYMPOSIUM
THE NATURE OF METAPHYSICS
THE REVOLUTION IN PHILOSOPHY

FREEDOM AND THE WILL

EDITED BY

D. F. PEARS

Student of Christ Church
Oxford

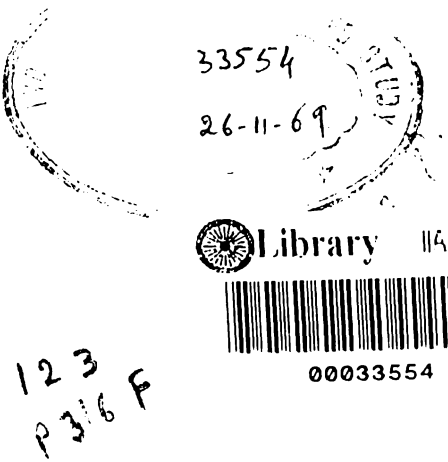
MACMILLAN
London · Melbourne · Toronto

ST MARTIN'S PRESS
New York
1969

First Edition 1963
Reprinted 1964, 1965, 1966, 1969

Published by
MACMILLAN AND CO LTD
Little Essex Street London WC2
and also at Bombay Calcutta and Madras
Macmillan South Africa (Publishers) Pty Ltd Johannesburg
The Macmillan Company of Australia Pty Ltd Melbourne
The Macmillan Company of Canada Ltd Toronto
St Martin's Press Inc New York
Gill and Macmillan Ltd Dublin

Printed in Great Britain by
LOWE AND BRYDONE (PRINTERS) LTD
London



CONTENTS

PAGE

1. FREEDOM AND THE WILL 1

By B. A. O. WILLIAMS, Knightbridge Professor of Moral Philosophy, Cambridge.

2. WHAT IS THE WILL? 14

By D. F. PEARS, Student of Christ Church, Oxford,
J. F. THOMSON, Professor of Philosophy, Massachusetts Institute of Technology, Cambridge, Massachusetts.
MARY WARNOCK

3. ACTS OF WILL AND LEGAL RESPONSIBILITY 38

By H. L. A. HART, Professor of Jurisprudence in the University of Oxford.

4. DETERMINISM 48

By G. J. WARNOCK, Fellow of Magdalen College, Oxford,
P. F. STRAWSON, Waynflete Professor of Metaphysical Philosophy, Oxford,
J. F. THOMSON, Professor of Philosophy, Massachusetts Institute of Technology, Cambridge, Massachusetts.

5. ACTIONS AND EVENTS 69

By G. J. WARNOCK, Fellow of Magdalen College, Oxford,

6. FREEDOM AND KNOWLEDGE 80

By IRIS MURDOCH
S. N. HAMPSHIRE, Professor of Philosophy, Princeton University, Princeton, New Jersey.
P. L. GARDINER, Fellow of Magdalen College, Oxford,
D. F. PEARS, Student of Christ Church, Oxford.

7. POSTSCRIPT 105

By B. A. O. WILLIAMS, Knightbridge Professor of Moral Philosophy, Cambridge.

Most of the essays in this book originated as talks in the Third Programme of the B.B.C.

1

FREEDOM AND THE WILL

We can trace back to near the beginning of Western Civilization the notion of individual responsibility: the notion that certain states of affairs in the world can be traced to the actions of human beings, for which those human beings can in various ways be called to account. The actions for which we are in the fullest sense responsible are those for which we can be praised or blamed.

This notion of responsibility is to be found already in the moral and legal thought of the Ancient Greeks. The Greeks already worked with the notion which we have today, that the question whether a man is responsible for something that happens is not *just* the question whether what happens is a consequence of movements of the man's body. For instance, suppose a man's arm moves in such a way as to knock over and break a valuable vase, someone else's property. In a limited sense, we know already in these circumstances what the cause of the damage was — the movement of this man's arm; but we do not yet know whether the man himself is in the full sense responsible for this damage, whether he is to blame. For that, we want to know more about the movement of his arm. In particular, we want to know whether he intended this movement of his arm, or whether perhaps it was just some nervous twitch, out of his

control. Again, even if the movement was not out of his control, we still want to know, for instance, whether he realized the vase was there: if not, his breaking of the vase will be unintentional, and to that extent free from blame. There are other sorts of situation, too, that relieve people of responsibility for things that they have done, or at least mitigate it; being in certain peculiar states, for instance, such as sleep-walking or under the influence of drugs; or, rather differently, being forced to do things by other persons.

Aristotle, in his *Ethics*, reviews and classifies these sorts of situation that relieve people of responsibility. His account is complex and subtle, but substantially he reduces the types of situation to two classes: those in which the agent is ignorant of relevant matters, and those in which, as he puts it, the originating principle of the action lies outside the agent himself — by which, I think, he principally means cases in which someone is physically forced to do something by someone else.

Aristotle's account is of interest for two reasons in particular. The first is a positive reason: that it is interesting to find a Greek philosopher giving an analysis of responsibility and the conditions that relieve people of it that so remarkably corresponds, in its essentials, to some of the considerations that we still employ today, both in the law and in everyday life. The second reason is a more negative one. Aristotle's account is concerned, as I have said, with the conditions that relieve people of responsibility for what they have done. I mean by this that he considers and classifies certain *special* circumstances

that relieve people of responsibility, while taking it for granted that in the usual circumstances people are responsible. He analyses the framework of praise and blame and responsibility as a going concern, as it were. He never, or scarcely ever, considers the notion that we might *never* really be responsible, that the going concern of praise and blame, our ordinary notions on these matters, might as a whole be founded on an illusion. Aristotle, that is to say, is precisely not concerned with that large-scale philosophical problem or set of philosophical problems which in later times has come to be known as the problem of freedom of the will. For the heart of that problem lies in the fundamental and revolutionary suggestion that our ordinary notions of responsibility may be altogether confused, because based on some false or ultimately unintelligible theory about human beings and their actions.

I think we may be able to see, in historical terms, how this fundamental and revolutionary suggestion came to be made, if we ask first why Aristotle does not consider it. One reason is that Aristotle regards it as certain, and indeed makes it a central point of his philosophy, that there is no necessity about human actions — that it is never necessary in any sense that a man should, on a particular occasion, have done this rather than that. For him, necessity is something that applies only to such things as the movements of the heavenly bodies; human actions, on the contrary, are a sort of thing which could always have happened otherwise. Now Aristotle did believe, I think, that *if* it could be shown that it was a matter of necessity that men acted in one way rather than another, then

our ordinary thought about action and responsibility would be undermined. If human actions were necessitated, there would indeed be something radically wrong with our notions of human action; but, he thought, it was quite certainly false that human actions were necessitated. If this is a correct account of Aristotle, we can see one reason why he does not confront the freewill problem. It is because he thinks that it cannot seriously be doubted that human actions are free from necessitation.

This is a very important point, because, as we shall see, one reason that the freewill problem did eventually arise was that men did begin to have serious doubts about just this. I shall come back to this point. First, however, it is worth while, for our historical picture, to look briefly at another reason for Aristotle's silence on this subject. This was his lack of belief in any personal god concerned with human affairs. Some earlier Greek writers, in particular the tragedian Aeschylus, do seem to be concerned with problems not far removed from the problem of freewill. Aeschylus's portrayal of Prometheus, or again of Orestes, seems to be in part a dramatic representation of human freedom as against forces set in motion by the gods, or perhaps we should better say, personified in the gods. Plato, nearly a century later, could still, in a poetic passage, write in these terms, and assert human freedom: 'the responsibility is with the soul that chooses its destiny: God is not responsible'. For Aristotle, a little later, there are no such gods, there are no such forces, and the question does not arise.

This is worth mentioning, because the problem of

freewill makes its first large-scale appearance in a religious context, when men had come to believe that there was one God, omnipotent, omniscient, and concerned with human action. The problem of freewill was first definitely stated as a problem of Christian theology. The problem arose, in fact, from a number of different roots in Christian belief: Christianity asserts on the one hand that man does freely choose his actions, but also asserts on the other hand statements not evidently compatible with this, for instance that God being omniscient knows from all eternity what actions a man will in fact perform.

I shall not say anything more about the theological forms of the freewill problem, nor will they figure in the other essays in this book. The theological forms are, not surprisingly, bound up with issues that are both special to Christian belief, and wider than the freewill problem itself; thus any theological discussion of the problem must involve the wider theological question of what it means to say that God's knowledge is outside time, and the special dogmatic issue of the operation of Grace, around which much of the Christian controversy about freewill has historically centred. The freewill problem, however, reaches us today in other forms not so theologically encumbered; and it is to these that the topics discussed in the present book will be specially relevant.

I mentioned earlier that men came eventually to doubt Aristotle's principle that human actions were not necessitated. This they did — or at least, the doubt occurred to them — with the rise of a mechanistic view of the universe, according to which the

universe was a closed system, every state of which was determined as a consequence of its earlier states in accordance with natural laws — laws which, in the eighteenth and nineteenth centuries, were believed to have been in essence discovered by Newton. The possibility of such a scientific account of the universe had occurred, indeed, to certain of the Greeks, who realized further that such an account would have consequences for human action. The atomist Epicurus, for instance, was apparently aware of these problems. The speculations of the atomists, however, remained speculations; and it was only with the rise of a successful experimental and predictive mechanical science in the seventeenth century that the problems became acute. How, it was asked, could a human agent intervene in the world, be, in Aristotle's terms, an 'originating principle of action', if every event in the universe was as it was merely in virtue of the earlier states of the universe, however remote? Would not every human action be a case of 'the originating principle lying outside the agent himself', so that there would be no action left for which he would be responsible? Descartes, in the first half of the seventeenth century, both held the mechanistic view of the material universe, and saw this problem; he wished further to safeguard responsible action from the realm of natural necessity. This he tried to do by distinguishing two quite separate realms, that of mind and that of matter, to the second of which alone, matter, the natural laws applied. But this was a quite inadequate kind of solution, even in Descartes's terms. Descartes himself realized that a bodily human action is itself, from

one point of view, an event in the material world. If all events in the material world are subject to natural law, how does the mind come into it? Descartes' distinction between two realms, designed to insulate responsible human action from mechanical causation, insulated the world of mechanical causation, that is to say, the whole of the external world, from responsible human action. Man would be free only if there was nothing he could do. Once this way of looking at the problem is accepted, there seem to be only two possibilities. Either all physical changes are subject to natural law, in which case those particular physical changes which constitute human actions are subject to it; or human actions are not subject to natural law, so not all physical changes are subject to it. The belief that all physical changes are subject to natural law has attracted, needless to say, a label: the label 'determinism'.[1] So here we have in outline an apparently simple opposition: either determinism is true, in which case there is no genuine human intervention in the world, that is, no freewill; or there is freewill, in which case determinism is false.

The issue has, of course, been endlessly posed in these terms, and in these terms endlessly discussed. But not all thinkers have accepted these terms, either explicitly by rallying to one side or the other, or implicitly by professing ignorance as to which alternative was true. Right from the seventeenth century there have been philosophers who have

[1] Strictly speaking, this is the notion of *physical* determinism. There are also ideas of psychological determinism: see Chapter 4; and Postscript, pp. 106 *et seq.*

criticized this opposition itself, and the terms in which the question was posed. Some have claimed, radically, that when the notions involved are properly understood, the alleged opposition dissolves, and the supposed dilemma can be shown to be a set of muddles and misunderstandings. Hobbes, Hume, and in our own time A. J. Ayer and others have taken this view. Others, less ambitious, have held that although there may be one, or probably more than one, genuine opposition here, nevertheless the terms of the discussion are so imprecise and ill understood that, without much further analysis, it is impossible to understand what the oppositions are, if they exist. Such a view is, I think, held by many linguistic philosophers today.

It certainly does seem that a tremendous number of questions have been begged and important distinctions blurred in the simple sort of formulation of the problem that I sketched just now. First of all, the formulation introduced a good deal of metaphor: one spoke of physical changes being 'subject to' natural laws, as though the natural laws exercised some mechanistic tyranny over events. Though an enticing metaphor, it is a very misleading one, and should be eliminated. What then, without benefit of metaphor, will be the formulation of determinism? It would seem to be something like this: that given any total state-description of the universe, it is in principle possible to predict or retrodict correctly any other, however remote, in virtue of a finite set of scientific laws. But there are still difficulties about this formulation. It is still not free from unclarities and ambiguities — great difficulties, for instance,

surround that slippery phrase 'in principle'. Under what conditions could we say that we could *in principle* predict every state of the universe? Just when we *thought* that there were laws in virtue of which we could do it, if only we knew them? This would surely not be enough — this would only be a situation of faith in determinism, not a situation in which determinism had been shown to be true. So perhaps we should say that determinism had been shown to be true if we *knew* all the relevant laws, and could predict all the states of the universe if only we took enough trouble. But then what certainty would we have that we did know all the relevant laws, unless we could actually succeed in making these vast predictions — that is, could not only in principle, but in practice predict? And this even the most fervent determinist would surely agree we could not do.

There are other difficulties in the formulation of determinism. What, for instance, are we to make of the phrase 'a total state-description of the universe'? Even if we made more precise — as we should have to — what terms such a description would have to be in, are there not overwhelming reasons for thinking that no such description could ever be completed? In the eighteenth century the astronomer Laplace could perhaps talk glibly in these terms; in the twentieth we certainly cannot.

Here someone may say: Good. We now understand that the conditions specified by the determinist could never be satisfied — that is to say, determinism is a false or incoherent doctrine. So the enemy of freewill is out of business, and freewill is all right. But such a confident answer would be premature,

and this just illustrates the obscurities of the simple opposition we originally set up. For while this extremely grandiose enemy, Laplacean determinism, is perhaps out of business, rather humbler but more effective enemies are certainly still in existence.

Here it is worth while to recall the point we started with right at the beginning: the fact that in ordinary life and the law we admit certain conditions as relieving an agent from responsibility for what he has done: conditions such as somnambulism, *force majeure*, etc. Now it is a most conspicuous feature of our present moral thought that this list of conditions is gradually being extended in the light of advancing psychological and other scientific knowledge. For instance, we now recognize in ordinary life, and perhaps even in the law, the existence of certain compulsive conditions — kleptomania is one sort of example. The proof that an agent was in such a state relieves him of responsibility, or at least mitigates it, for actions of the appropriate type.

Now it is not an accident that the discussion of such cases gives rise to constant difficulty and doubt. We feel compelled to admit more and more such conditions into the class of conditions that exonerate, without really being clear on what principle we are doing it. Moreover, there is the lurking feeling that the principle on which we are doing it might be one that eventually might extend to swallow up wide ranges of action now regarded as normal and responsible. For instance, are we to say that a man is not responsible if there is a psychological explanation of what he did? This by itself is too weak, since the phrase 'psychological explanation' can cover

practically anything, including for instance 'he freely chose'. But perhaps some sorts of psychological explanation . . .? Here we do not know, not just *where* to draw the line, but *how* to, and in this ignorance we can feel no *a priori* confidence how much of our ordinary sphere of action will be left intact by the line when properly drawn. Here we feel the presence perhaps of another type of determinism less grandiose but more pressing than the total physical determinism of Laplace.

In this situation, looking for a criterion, we may turn to the other side of the field. We may ask, not what are the general conditions of non-responsibility, but what are the general conditions of responsibility: perhaps there is some sign that an action is, in the appropriate sense, really ours. Here we meet for the first time a concept that has been absent from the discussion so far — the concept of the will itself. It is in virtue of the operation of the will that some philosophers have tried to distinguish this class of responsible actions. But what is the operation of the will, and what are its signs? We indeed speak in ordinary life of 'efforts of will', and it is in the occurrence of these that in reflective moments we perhaps feel most conscious in some sense of our freedom. But here there are many difficulties. First, it is certain that only a very few of the actions for which people are normally held responsible are accompanied by efforts of will, in this psychological sense; and the same goes for any other conscious process that might be suggested here instead, such as explicit decision, formulated intention and so on. Often we just act, without such processes, nor would we

regard such actions as any the less responsible or free for that reason. Again, there is a deep difficulty about what an effort of will really is. There is indeed some kind of psychological process in connection with which the term is used — but might it not be *just* a psychological process which accompanied some actions (perhaps peculiarly difficult ones) and not others? Contrary to what some philosophers have supposed, efforts of will do not wear their metaphysical significance on their face. It may even be that they have none.

I have mentioned in the last few pages a good number of problems and difficulties. It is these difficulties that contributors to this book will be considering: in particular, the definition of determinism; the search for general conditions of responsibility; the nature of the will and its connection with what we call efforts of will; the scope and implications of different kinds of psychological explanation. My aim has been to try to link with a few threads these different topics to the history of that set of issues which is over-simply called 'the problem of the freedom of the will'. I have tried to suggest that there is not one simple issue here, but a great network of difficulties. The difficulties are, however, not unrelated to one another — the network has a centre. This centre lies on a line, which is deeply engraved in human thought, the line between man as a conscious, reflective being and man as part of physical nature, conditioned by and acted upon by his environment. The basic nature of this dividing line appears in the constant preoccupation of philosophy with these difficulties; and it is its very depth

that makes it necessary for any serious search for understanding in this region now to distrust grandiose theories, and to take the difficulties, so far as possible, one at a time. They are, after all, no less complex than the human situation.

2

WHAT IS THE WILL?

DAVID PEARS. The phrase 'the freedom of the will' occurs in many philosophical discussions. But if we examine what goes on in these discussions, we find philosophers talking about predictability, about motives, about the various ways there are of explaining human behaviour, about freedom and its meanings; and we do not find them talking very much about the will itself. They do not ask what it is to which we are or are not to ascribe freedom. Is this a remarkable oversight, or is there some other explanation?

J. F. THOMSON. Need the phrase 'the freedom of the will' be taken seriously? Don't we nowadays use it as a conveniently vague pointer with which to gesture towards a whole set of philosophical issues?

MARY WARNOCK. Yes, I think we do. But we can still ask what the will is: we can take that seriously. For instance, we sometimes say that someone did something of his own free will. That is not a piece of philosopher's jargon, but a phrase which everyone uses and understands. And the mere existence of this phrase is a *prima facie* reason for thinking that there is something to ask about.

J. F. T. If I know what is meant by saying that someone did something of his own free will, it doesn't follow that I have to admit that he has a

will. The explanation of the phrase 'of his own free will' needn't involve something called 'the will'.

M. W. But I could just as well say that if you know he did it of his own free will, you *ipso facto* know that on this occasion his will was free.

J. F. T. Well, if you tell me that I am to say that his will was free, if, and only if, he did whatever it was of his own free will, then of course I now understand this phrase about his will. But it doesn't help me to understand any more than that: I can't see that anything has really been accomplished.

M. W. Perhaps not much yet. But there are other phrases in common use which seem to make reference to the will: we speak of people being strong willed or weak willed.

D. F. P. And there are efforts of will and will power.

M. W. There is a set of recognizable phenomena which, roughly speaking, involve the will. What they have in common is their relevance to situations of initiating and carrying on action and getting things done. To investigate these would be to investigate the will.

J. F. T. Perhaps I'm being over cautious. I don't want to deny that there is something to look into. But I do still want to reject the idea of the will as a special ability or power to translate thoughts and intentions into actions. This conception of the will is implicit in common speech and explicit in a lot of earlier psychology—*e.g.* Locke, Hume, Reid—and it seems to me a useless one. I don't deny, of course, that we do 'translate intentions into action',

nor that we sometimes exercise strength of will in doing so. What I do deny is that there is any call to postulate a special unique ability to do so. You might say that when I raise my hand, my hand has obeyed my will. But nothing is gained by this 'proof' that the will exists; it is just a pointless verbal manœuvre.

D. F. P. I'm not sure why you call it that. Is it because you have a *general* objection to talking about faculties? If so, you would, I suppose, be equally averse from talking about the intellect? Yet this could be a quite innocuous way of referring to intellectual activities. Or is it rather that there are no activities which are *just* activities of the will, as there are activities which are just intellectual or just physical? If this is your reason, perhaps we ought to examine the attempts that some philosophers have made to identify an activity of willing. For, of course, the difficulty is that the activity of willing is supposed to be involved in every activity.

J. F. T. Yes, that is the difficulty, and I think it is one that is very hard to overcome. Let me explain why I spoke of a pointless verbal manœuvre. Suppose I can say truly 'I raised my hand'; as distinct from, merely, 'my hand went up'. And suppose a philosopher tells me that I must have willed my hand to go up. Then perhaps I shall understand him to mean that I can say 'I willed my hand to go up, when, and only when, I have raised my hand'. But if this is *all* I understand by, 'I willed my hand to go up', nothing has been accomplished, since 'I willed my hand to go up means *no more than* 'I raised my hand'. It is very obvious that the notion of willing

isn't doing any work at all. All we have done is to add a word to our notation. This difficulty can be met, but only at the cost of meeting a greater one. It may be said that my *willing* my hand to go up is not, as I supposed, meant to be a sufficient condition for my hand going up. In other words, the theory allows for the possibility of acts of will misfiring or being somehow ineffective. The question must then be faced, how are we supposed to tell when an act of will has genuinely occurred, but been ineffective? This question must surely prove embarrassing. But, until it is answered, it can't be thought that the idea of acts of will has been properly explained. So I think that anyone who wants to talk of acts of will can be faced with a dilemma; either the concept of willing is trivial and parasitic, or it is as yet quite unexplained and obscure.

D. F. P. Yes, I think that is a fair dilemma. And it could be used to divide theories of the will into two classes. Some philosophers argue on *a priori* grounds that there must be such an activity as willing or that there must be acts of will, but do not try to identify this activity empirically. Others do attempt to identify it empirically, at least in some cases.

M. W. Could we take the second kind of theory first, and try to see whether acts of will *can* be identified empirically? Why shouldn't we simply say that acts of will are the efforts of will we spoke of before? Suppose that I am terribly tired and sleepy, and have two more urgent letters to write before I go to bed. To make it more poignant, imagine I'm in my bed-sitting-room, even sitting on the bed, and all I have to do is to lie down and go to sleep. Now

I can argue with myself that I have *got* to get these letters done, I can tell myself to keep awake, and by the most tremendous effort I can resist lying down to sleep until I have finished the letters. It might well be said that the difference between the person in my situation who gives up and goes to sleep, and the person who finishes the letters first, is that the virtuous one finished by an act of will. And there are lots of cases like this one — the man who wants a drink and doesn't have one, and so on.

J. F. T. But do you mean that acts of will just are efforts of will, and conversely? If so, then since I've agreed that there are efforts of will, I've implicitly agreed that there are acts of will too. If we look at the matter in this way, it is certain — and not merely arguable — that there are acts of will. But if this is all that is meant by saying that there are acts of will, nothing is explained by it. But perhaps you were suggesting that acts of will occur whenever efforts of will occur, although the two things are not identical. If so, my difficulty is really the same as before; when I make an effort of will, I can detect nothing in my experience, over and above the effort, that I would want to call an act of will, and I simply see no reason to suppose there must be one.

D. F. P. Isn't it true anyway that we speak of having made an effort of will only when we have been in some kind of conflict? Take your example: if you never wavered for a moment in your determination to write the letters before going to sleep, then, however tired you had been at the time, you wouldn't, I think, say that you had made an effort

of will, and this because you wouldn't have *needed* to make one. But, as you describe it, you are in a position in which it would be very easy — all too easy — to go to sleep right away. For this reason the thought of going to sleep is one that very easily occurs to you. And because it occurs to you, it has to be *put on one side.* I would like to say that the effort of will consists in putting it on one side, in continually turning your attention from pillow to page. But if this is right, then making an effort of will is doing something which is perfectly familiar, and something which is noticeable when it is done; and it is certainly not something that has to be done before every overt action.

M. W. I certainly don't want to say that there has to be an effort of will before we do anything. That would just be false. And because it's false we don't have to be afraid of the objection that is often made against theories of the will, that they lead to an infinite regress. It is often argued — for instance by Professor Ryle in the Concept of Mind — that if we say that every action must be preceded by an act of will, then, since acts of will are themselves actions, it follows that every act of will must be preceded by another act of will, and so on *ad infinitum.* If this infinite regress did arise it would indeed be a vicious one. But we can easily evade the whole thing by not requiring that *every* action be preceded by an act of will. And it seems to be a simple empirical fact that some are and some are not.

J. F. T. Well, you are now implying that an act of will just is what we ordinarily call an effort of

will. And, as I said before, this means that the phrase 'act of will' adds nothing to our knowledge. There is no harm in using it as long as we remember that acts of will do not precede every action. But I suspect that just for this reason, most philosophers and psychologists who have wanted to theorize about the will would be dissatisfied with your account. Its very sobriety would damn it in their eyes. For what they wanted to establish is that the will is essentially involved in every voluntary action.

D. F. P. I share your suspicion. But if that is what has to be established, it is difficult to see how it could be established empirically. Certainly not by appealing to efforts of will. But perhaps something might be done with the help of a different line of argument. Suppose that a man's right arm is suddenly paralysed without his knowing it. I say to him: 'Could you raise your right arm now if you wanted to?' He says: 'Yes, of course.' Then I ask him to do so, and, to his amazement, his arm remains at his side. A philosopher who wanted to say that acts of will could be identified empirically might point to a case like this and say that here, surely, an act of will must have been performed.

J. F. T. Because, he would say, why otherwise should the man be surprised at his arm not going up?

M. W. Yes, we have to agree that he tried to raise his arm, however hard it may be to say what this trying consisted in.

J. F. T. But let's not lose sight of that difficulty —just the question at issue is whether his trying constituted an act of will.

D. F. P. But I feel like asking 'What else could

it be?' For no physical movement or change occurs in his body.

J. F. T. I see the force of that, but I'm not convinced. For one thing, how do we *know* that no physical change occurs in the man's body? Why shouldn't it be that he does succeed in moving some muscles, but not all of the right ones, or that he succeeds in bringing about a contraction, but not a sufficiently strong one? Second, why should we take it for granted that, whenever someone tries to do something, even if he fails, he always succeeds in doing something else, and that this is what his trying consists in?

D. F. P. Yes, and now that I think about it, are we even right in assuming that this man *tried* to raise his arm? If a victim of poliomyelitis undergoes a course of treatment intended to give him back the use of his legs, he might be told at some stage to try to move them, and here the word 'try' does seem in place. But the case I described is rather different, because in that case there is neither the same kind of effort nor the idea that there might be failure. However, perhaps this isn't as important as one of the assumptions that James Thomson questioned — that whenever someone tries to do something, then, even if he fails, there is something which he does succeed in doing, and it is this that his trying consists in. Now the reason why I was making this assumption is that, if he hadn't *done* anything, there would be no reason for him to feel surprised when his arm remained at his side. I admit that we don't really know that there wasn't some change in his body. But isn't it plausible to say that, even if there were

no change, still the man must have done something, since, otherwise, he would feel no surprise at the immobility of his arm?

M. W. Well, we don't really know that if there were no change in his body he would still feel surprised, do we? And anyway it seems to me that you aren't any longer doing what you set out to do. You set out to identify acts of will empirically. But now you are relying on an *a priori* assumption that the man in your example must have done something, and you haven't given any account of what it is that he is supposed to have done. In fact I think your attempt belongs to the first of the two kinds of accounts you mentioned earlier on — that is an *a priori* account, and not, as you have been suggesting, the second, empirical kind.

D. F. P. Perhaps so. But I am still puzzled by this. How is the man's surprise to be explained?

J. F. T. Yes, in spite of what we've said, this question is still oddly obsessive. The question really is, how does he *discover* that he can't raise his arm? It may well be that the right answer to that question is that it depends on the particular case. But here we're handicapped by our ignorance of what actually happens, and what is physiologically possible. But I think that what really makes the question difficult is that we approach it with a concealed assumption. This is that the man discovers that he cannot raise his arm by *doing* something — something quite unspecified — and then discovering that this thing which he does, which used to be followed by his arm going up, is no longer followed by that happening. (Of course in most cases of unsuccessful

trying, this is just what does happen.) I try to stop my car by putting on the brakes; I fail because the master cylinder is empty; but I do succeed in depressing the brake pedal. And here it is, just because I have succeeded in depressing the pedal, that I learn that doing this does not stop the car. I think your question is worrying because we take it for granted that all cases of unsuccessful trying have to be assimilated to this one; but this is a mere assumption.

M. W. I suppose there must *be* a reason for his surprise, in the sense of an explanation, but it doesn't follow that *he* must have a reason. In some cases the explanation might be that when he was asked to raise his arm, he was, to use a term psychologists use, *in a set* to obey and was aware of this. This doesn't mean that he had deliberately decided to raise his arm, nor that he had made any kind of effort or attempt to raise it. There are dozens of things that we do in this way — knowing that we are up to, but not planning or trying to do them. I may be surprised if I am frustrated in any of these cases. . . . I mean if I am typing, and my fingers are ready to go down on to the proper keys, I don't try to put them down or think about it at all, that is if typing comes pretty easily to me, like raising my arm. And I should be surprised in any such case, if what I was prepared to do didn't come off. But it would be absurd to identify this preparedness or expectation, or whatever we call it, with performing an act of will. That is why I said that this example doesn't supply us with an empirical reason for postulating acts of will. Explanations of the man's surprise, like the one I've just mentioned, seem to be tacitly

excluded and their exclusion rests on an *a priori* assumption.

J. F. T. Yes, and I think we can see from this example what the weakness of the *a priori* approach is. For suppose this man who is trying to raise his arm *does* do something, call it *X*, which, because his arm is paralysed, isn't followed by his arm going up; then surely it is conceivable that he should be unable to do *X*, and the same question would arise all over again about how he discovered that he couldn't do that.

D. F. P. I should like to generalize that objection. What you called a concealed assumption could be called a *conceptual demand*; the demand that whenever someone does something, there must always be some second thing he does as a means to doing it. And the objection in its most general form is that the demand must then be made about the second thing that he does as well — and so on *ad infinitum*. So the theory ends up by saying that all action is indirect. And your objection is that no action can be indirect unless some is direct.

J. F. T. But this isn't all that's wrong with the *a priori* approach. Even when there is a second thing that happens, it won't always be done by the agent as a means to doing the first thing. I depress the brake pedal in order to put on the brakes. But would it be correct to say that in order to depress the brake pedal I contract some muscles in my calf? It's not obvious that it would. And to go a stage further, it is plainly false to say that in order to contract the calf muscles, I send off an electrical impulse from the motor cortex. A human body just

isn't like a machine with controls which have to be identified and whose function has to be learned.

M. W. But I'm afraid I have a difficulty about the main objection — the one which turns on an infinite regress. I sometimes wonder whether such arguments are as strong as they look. Perhaps in this case the objection itself rests on illegitimate assumptions.

J. F. T. Well, the objection certainly doesn't rest on the assumption that every act needs an anterior act. Precisely the reverse. The *a priori* thesis rests on that assumption and the objection is attacking it. Earlier on you pointed out that someone might hold that acts of will are efforts of will, and you said that this thesis was not vulnerable to the objection that it led to an infinite regress. That was quite right, because your thesis didn't then rest on the assumption that *every* act needs an anterior act, but relied on an empirical or phenomenological identification of acts of will in some but not all cases. But the *a priori* thesis *is* based on the assumption that every act needs an anterior act, and consequently it *is* open to the objection that it leads to an infinite regress.

M. W. But isn't it possible — just possible, to reply that acts of will are different from other acts precisely in this respect, that they don't need anterior acts?

J. F. T. That wouldn't do. The whole force of the *a priori* argument — I mean what force it has — is derived from the analogy with ordinary acts, like putting on the brakes and depressing the brake

pedal. If you abandon this analogy and talk about acts of a different kind, the argument collapses at once. Briefly, if some acts — acts of will — don't need anterior acts, why should all overt acts have to have them? Once you make an exception in favour of acts of will, you are giving up the general assumption, and then the question whether a given overt act has an anterior act has to be answered on the merits of that particular case.

D. F. P. Perhaps the most fundamental error in the *a priori* thesis is that it assumes that the connection between two events can be explained only by postulating a third one. For in the typical case the agent begins by thinking of his project and planning it, and then eventually he executes it. Then the question is, 'What is the connection between these two events, the thinking and the execution?' Now sometimes there are events between these two: for instance, he makes up his mind to perform the action, or he makes a special effort to perform it. But it is a mere assumption that there must always be intervening events. And in fact we just find ourselves doing many of the things that we do do.

J. F. T. And if the assumed intervening event is another action, nothing is explained anyway, for then the same question arises about the connection between the thinking and that action.

D. F. P. Yes, because the connection between the earliest action and the agent's thinking of it remains for ever as unexplained as the connection between the overt action and the thinking about that. But I think there is another problem here, about the connection between the act of will and the

overt action to which it is supposed to give rise. What is the nature of *this* connection meant to be? I mean, how could acts of will explain the actions that followed them. If they did explain them it would be because there was some connection between the two things, and we would have to understand the nature of the connection.

M. W. It occurs to me that there is a kind of willing which, if it happened, really would be explanatory of what it brought about. I mean psychokinesis. In some of Rhine's experiments, dice were thrown up and the subject was asked 'to will the dice to land with a specified face or combination of faces uppermost'. And Rhine certainly seems to think that in some of these experiments results were obtained which are most readily explained by supposing that the subject was able to influence the dice to fall in the way he instructed.

D. F. P. Yes, and this kind of willing might throw some light on the ordinary concept of willing. The connection between this kind of willing and the result might be claimed to be like the connection between an act of will and an action.

M. W. Well, let's work at this kind of willing to see if it does throw any light on the so-called ordinary concept. But of course we mustn't forget that, as we seem to have shown already, the ordinary concept is highly problematic.

J. F. T. Well it's clear that whether or not Rhine's hypothesis is correct, it is perfectly intelligible. The suggestion is that someone has a certain power over physical objects, and the dice fall the way they do because he chooses to exercise that power. We

can easily imagine a situation in which we would be practically forced to admit that someone does have such a power. He announces that he is going to will something to move — a book, say, which is at a distance from him; then he sits perfectly still, perhaps with a look of concentration, and the book moves. Naturally we should at first suspect all kinds of conjuring tricks. But we can imagine too that our suspicions of this kind get allayed. Then it would be reasonable to say that this man can move things just by willing them to move.

M. W. I agree that in one way it's quite conceivable that this sort of thing should happen. I mean one can understand what's meant by the description of it. But you're not asking us merely to imagine that a book rises up into the air. We've got to imagine also that this happens *because* someone *wills* it to. So haven't you got to be able to say just what his willing the book to move consists in, — how he does it?

J. F. T. I don't *think* so. Or rather I don't think these two questions are the same. If we ask him how he wills the book to move he may say, 'I don't know. I just do it.' Consider: you find by accident that you can serve in table tennis in such a way that the ball doesn't bounce on the other side of the table but runs along it. That you can do this is shown by sufficiently frequent success. But you may be quite unable to see *how* you do it. I think we would say we know how you did it only if we had some idea of what to do ourselves. And the same in the willing case. The question of what his willing consists in is rather different. That is a question of the outcome

and not of the means, method, technique or *modus operandi*. His willing the book to move consists in the book's behaving in the way we described and in its doing that because, in some perhaps unknown way, he makes it do so.

M. W. So then 'he wills it to move' *means* that he makes things move but without touching them or using physical instruments? In your sense, willing things to move is a way of making them move the way that kicking them is a way of making them move. Willing that something should happen is a way of making it happen.

J. F. T. Yes, in the sense that we have explained it, it is. Willing the book to move is moving it *by willing*.

D. F. P. But then it turns out, doesn't it, that the question 'can we will things at a distance from us to move?' is ambiguous. Suppose someone says that he can will things to move. We set him a quite easy test, and ask him to will a pencil to roll off the table. Nothing happens. 'But', he says, 'I only claimed I could will things to move. I didn't say they would move when I willed them to.' This man now can't be using the phrase 'will things to move' in the way you explained; if when he wills things to move and they don't move, he isn't at all surprised, he can't be claiming to be able to move things by willing.

J. F. T. No, that's quite right. But how are these two notions of willing related? or are there two notions?

D. F. P. It's not obvious that there are, is it? If someone said: 'I certainly don't believe I can will

things to move in your sense — I don't believe I can move them by willing. But all the same I can will them to move in a weaker sense. I can will that book to move in just the same sense as I can will to get up on a cold morning. The only difference is, that when I do the latter I do sometimes get out of bed, whereas my willing the book to move is never followed by the book's moving.' Wouldn't this somehow seem an absurd thing to say?

M. W. I think it would, but I think we can see why. We should ask him what he did when he willed the book to move — what his willing consisted in. The answer 'making the book move' isn't now available to him, because the book isn't actually made to move. But some answer must be forthcoming. So suppose he says something about staring at the book and imagining its moving. Then we can ask, why should this be called 'willing the book to move'? If staring is given this title just by a linguistic fiat, the case ceases to be interesting, because then it's clear that he's saying only that he can stare at books and what's so interesting about that? 'Willing the book to move' in this weaker sense is interesting only if it means 'trying to move the book by willing', that is, trying to *will to move it* in the earlier sense. But now the absurdity becomes manifest. If you say that by doing X you are trying to do Y, you imply that you think there's some chance that by doing X you'll succeed in doing Y. So if you say that your willing the book to move consists in staring at it you imply a belief that staring at things at least sometimes makes them move. Of course if you do have this belief, then your statement that you can will things

to move is, so to speak, conceptually correct, but your belief is false. If, on the other hand, you say that you can will things to move while disavowing any belief in the possibility of moving them by willing then you are just confused.

D. F. P. Yes, we were imagining that someone *now*, as things are, said he could will things to move in an alleged weaker sense; like William James or Prichard saying that he could will the distant table to slide towards him as easily as he could will to write. If someone who had already shown that he could make tables slide towards him just by willing them to were to say: 'I am trying to will that table to slide towards me, but I can't somehow seem to manage it', the case would be altogether different. *This* man has the right to say 'I am now doing whatever it is that I do when I succeed in willing tables to slide towards me, or at least I seem to myself to be doing it, but the table doesn't move all the same'.

J. F. T. It seems then that we haven't got two senses of willing here. There is willing things to move in the sense of making them move by willing, and there is trying to will them to move in that sense. As things are, we can't will things to move, and, largely because of this, we can't try to will them to move either.

M. W. Mostly I agree with this conclusion. I think it's true and important that the special kind of willing need not always involve an identifiable performance which could be explained to someone else; and that if there is an identifiable performance which can properly be called 'trying to will in the special way' it must be something which the person

himself believes to have a reasonable chance of success. But there are two qualifications which I should like to add, both about trying. The first is that when there is a performance which the person himself believes to have a reasonable chance of success, it need not be called 'trying', for it might be immediately successful, and then it would be called 'his method of willing'. The second qualification arises out of what David Pears said. It is that, when he does try to will, there may not have to be an identifiable performance. For, as we have already agreed, in perfectly ordinary cases where a person tries to do something, his trying need not consist in any identifiable performance. And so it seems reasonable to allow that, by analogy, trying to will in the special way need not consist in an identifiable performance. I suggest that we are apt to assume that there must be an identifiable performance only because the special kind of willing does not, as a matter of fact, exist, or at least is not prevalent; and so when we think about it, we feel bound to postulate identifiable performances to fill the void. First, we postulate that there is a performance every time anyone wills something in the special way, and then we postulate that there is a performance every time anyone tries to will something in the special way. But the performances are unnecessary in both cases.

J. F. T. If this is right, it has an important bearing on our main problem. For we set out to determine whether the special kind of willing throws any light on ordinary acts of will, like the act of will that sometimes occurs when one gets out of bed; and, in particular, we wanted to discover whether

the connection between the special kind of willing and its result would throw any light on the connection between the act of will and the action. And we can now see that it doesn't throw any light at all. For the account that we have given of the special kind of willing is simply based at every point on the account that we would give of ordinary actions. So no wonder the two accounts are parallel. It is ordinary actions that are throwing light on the special kind of willing.

D. F. P. Is it quite as bad as you say? I think we do learn something by examining the special kind of willing. For suppose we consider the cases where no identifiable performance is involved, then don't we discover something really rather interesting, that in these cases there is no problem about the connection between the special kind of willing and the result, simply because there are not two identifiable things to be connected? Of course this will probably lead people to object that it makes it impossible to see how anyone would ever learn to will that the table should slide towards him. But the objection is invalid. For it just assumes that learning how to will this must be learning how to execute some identifiable performance. But why should it be? When a child learns how to raise his right arm, he doesn't learn how to execute some identifiable performance — except raising his right arm.

J. F. T. But we could have reached this conclusion, and did reach it, without going into the question of the special kind of willing.

D. F. P. Perhaps. But it is salutary to realize

that, even with the special kind of willing where we are very strongly inclined to postulate an identifiable performance, it still isn't necessary. This is salutary because the inclination is a general one, and most misleading. James Thomson alluded to it earlier, when he said that we tend to regard the human body as a machine with controls which have to be identified, and whose function has to be learned. Of course we do not learn our way around our muscles and nerves in this way. And similarly we don't learn to discriminate between different acts of will in this way. The operation of the will is not like the movement of the innermost set of levers. We identify what we have to do through the description of the act itself. And this would often be so even with the special kind of willing.

M. W. But what about the other kind of cases, where the special kind of willing would involve an identifiable performance? Would these cases throw any light on the connection between the act of will that sometimes occurs before getting out of bed and the action of getting out of bed?

D. F. P. That is a complicated question. First, I suppose that the person who wills in the special kind of way knows what he is willing, and that it's not like the kind of magic that can be operated unwittingly. For instance, he concentrates very hard on the idea that his horse is going to win.

M. W. If that is right, then in this type of case, the special kind of willing is not exactly analogous to the act of will that sometimes occurs before one gets up. For it doesn't seem to be just a contingent fact that this act of will involves concentrating on

the idea of getting up. In other words, the point that you made earlier in the discussion, that an effort of will consists in concentrating one's attention on the idea of the action is a necessary truth.

J. F. T. But can't you imagine people whose acts of will consisted in barely thinking of the action and then dismissing it from the mind? Rather like the way that we sometimes recall names that we have forgotten?

M. W. I find that hard to imagine. For the acts of will of such people would be too unlike ordinary decisions. They wouldn't really be acts of will, but only devices or tricks for getting themselves to do things.

D. F. P. Yes, that is surely right. And it is connected with a very important difference between the special kind of willing and the ordinary act of will. In the ordinary case we will *to* do something, whereas in the special kind of willing, we will *that* something should happen. Now willing to do something has to be something like an ordinary decision, in an intensified form. It can't be entirely different from an ordinary decision. It can't therefore consist in barely thinking of the action and then dismissing it. On the other hand, if we will that an event should happen, and we want to join this willing on to a particular performance, absolutely any performance would do. The performance need not even involve thinking about the event: all that is necessary is that the person himself should believe that it has a reasonable chance of being followed by the event. Admittedly it is natural to think of concentrating on the idea. But that is only because we are influenced

by the analogy with acts of will. In fact any kind of performance would do, since there is no need for it to be like deciding to do something.

M. W. Yes, and this could be applied to another of William James's examples: willing to sneeze. Sneezing is not an action in the full sense, but, because it counts as a remote relation of proper action, we can talk about willing *to* sneeze. However, since there's no such thing as deciding to sneeze, willing to sneeze could take any form at all. We might find that concentrating on the idea of sneezing inhibited the action, and that the most effective method was more oblique.

J. F. T. Well, where have we come out? We have been trying to discover whether the special kind of willing throws any light on the will in ordinary cases of action. And I think the result is negative. For at certain points the two concepts are not parallel to one another. And when they are parallel, this is simply because we have constructed the concept of willing in the special way on the analogy with ordinary cases of action. Sometimes, it is true, this construction brings out some feature of the ordinary cases particularly clearly. But we don't learn anything new from it.

M. W. That certainly *is* a very negative result. And actually I think the whole discussion has been pretty negative. We've spent the whole time saying what the will is not, and never what it is.

D. F. P. But it's important to clear away misconceptions. And in this subject it's so easy to assume that we know what the will is, and so can go straight ahead and try to decide whether it's free or not.

But the situation isn't really like that at all. For when we look at it more closely we see that there are many things that we do which might or might not be free. We act, we choose, we decide, we try and sometimes we make efforts of will. These are the real candidates for freedom. But what usually happens is that they are supplanted by a single candidate with no identity, but only a name — the will.

J. F. T. Yes, the freedom of the will is not the freedom of a single thing, and if we use a single name we are using it compendiously. . . . Of course this need not do any harm, but it often has. For it has made it look as if there must be a single thing. But, as we have seen, attempts to identify such a thing empirically usually give it far too restricted a field of operation: and attempts to prove its existence *a priori* always fail.

D. F. P. But we need not regard what we have been saying as simply prolegomena to the problem of the freedom of the will. Certainly freedom of the will is connected with very important questions of legal and moral responsibility. But, quite apart from these questions, the psychological preliminaries to action are interesting in themselves. What *does* happen before we act? How does the sequence of deliberation and decision really go? Where does choice fit in? All these are questions of the greatest interest. But they have often been blocked by an unrealistic and excessively simple theory of the will, and so this obstacle has to be removed.

3

ACTS OF WILL AND LEGAL RESPONSIBILITY

In the previous discussion in this series, the speakers considered the theory that in every voluntary action there is involved a distinguishable activity of willing or an act of will. According to this theory if I raise my arm, and if this is a voluntary action on my part, then at some anterior stage I must have willed something. On some variants of the theory what I must have willed is the raising of my arm, but on others it may only be the initiating, bodily movement or contraction of my muscles.

This theory, as the speakers last time showed, cannot be supported either by *a priori* reasoning or by an appeal to the facts of experience. It is of course true that in special cases such as those where we have strong temptations to resist, there are phenomena which it is natural to describe as an effort of the will. If we have to work when we want to sleep we may, as we say, *make* ourselves work by a variety of devices: by concentrating on the page; by averting our eyes from the pillow; or by uttering self-addressed adjurations. But these efforts of the will, which are familiar enough in such situations of temptation and conflict, are certainly not present in all voluntary actions. So the theory that they must be present, looks like a philosopher's fantasy; it

seems merely to be a dogma about the nature of human action which has been adopted in advance of a detailed scrutiny of the facts. It is the dogma that every voluntary action which we do is something which we have to initiate by doing something else — even if, in the last resort, this something else is an invisible and, indeed, an undetectable act of willing.

I agree with the criticisms of the theory which the speakers last time made along such lines. But in this talk I wish to show that the theory is something more than a philosopher's fantasy. In the first place, it is a legal theorist's as well as a philosopher's fantasy; for it has figured for at least a hundred years in the general accounts of criminal responsibility given by legal theorists both in this country and in America. Moreover, when we consider it in the setting of the concrete issues of responsibility with which lawyers deal, it is possible to see that the temptation to adopt this theory springs quite naturally from the effort to characterize certain ways in which our conduct may be defective or not under our control. And these things are very relevant to questions of responsibility and difficult to characterize in general terms.

For the legal theorist the temptation to talk of acts of will as involved in all voluntary action arises in the following way. In all advanced legal systems responsibility for serious crimes depends not only on proof that the accused person's outward movements and their consequences are those comprised in the definition of the act which the law forbids, such as killing or wounding other persons. Responsibility also depends on certain mental elements: usually

upon proof that the accused had knowledge of the circumstances and foresight of the consequences of his outward movements without which he could not be said to have intended to do harm. Precisely what knowledge and foresight is required in the case of any given crime is a technical matter into which I shall not enter here. But suppose one man (Smith) has shot and wounded another (Jones): these facts will not be sufficient to establish Smith's criminal responsibility for wounding Jones if, for example, Smith thought, on good grounds, that the gun which he was playfully pointing at Jones was unloaded, or if Jones stepped into the line of fire just as Smith, who had no reason to anticipate this, was firing at a bird. In such cases we should say that though Smith wounded Jones he did it owing to a mistake or accident and so did it unintentionally.

Now the knowledge of circumstances and foresight of consequences which Smith lacked but would have possessed had he intentionally wounded Jones are included in what Anglo-American lawyers call *mens rea*. Generally *mens rea* in this sense is necessary for criminal responsibility, though there are certain cases where negligence may be enough and also cases such as the offence of dangerous driving where responsibility is said to be '*strict*' and is not dependent on either negligence or intention to do what the law forbids.

For present purposes the important point is that though the mistakes and accidents of the kind which made Smith's action in wounding Jones unintentional are the most familiar forms of aberration or defect in human conduct which generally exclude criminal responsibility, there are other and still more

fundamental defects. When these are present the movements of the human body are more like the movements of an inanimate thing than the action of a person. Suppose Smith had been suddenly startled by a loud noise or stung by a bee and had jumped and involuntarily pressed the trigger of the gun with fatal results for Jones, or suppose a man unconscious in a fit of epilepsy hits out and hurts someone. In referring to such cases ordinary people might find themselves vaguely reluctant to speak of them as cases of action at all; they might deny that the person concerned really 'did' anything. Legal theorists have exploited these tendencies of ordinary speech and where human conduct is defective in such very fundamental ways they say that there is no 'act' at all. This is then generalized into the doctrine that for legal purposes an act is something more than a mere movement of the body: it must be willed. The practical importance of this doctrine is that even in cases of strict responsibility like dangerous driving where *mens rea* is not required, none the less, according to the theory in the books, the law still requires that there must be an act in the sense of something willed, for this is a minimum link between mind and body indispensable even in the strictest forms of responsibility.

This doctrine has only rarely been considered by the courts, and I am myself not convinced that the courts actually do accept this general doctrine. [1] They seem to me to dispose of the cases by quite a different

[1] For a discussion of this point see the author's *Acts of Will and Responsibility* in the Jubilee Lectures of the Faculty of Law, University of Sheffield (London 1960).

technique. None the less you will find this doctrine still confidently, if a little automatically, repeated in most English textbooks on criminal law. Here are some of the stock phrases used. There must, it is said, be 'an act due to the deliberate exercise of the will' or 'voluntary conduct' or 'an act with its element of will'. But of course these are dark phrases. What, after all, is meant by the 'will' and its 'exercise'? The books give very little explicit guidance but instead provide a list of examples where there is said to be no real 'act' because there is no exercise of the will. Some of the examples are hypothetical and a bit far-fetched, but they may be divided into cases where the person concerned is conscious and those where he is unconscious. The conscious cases include the following. A man, *A*, is holding a knife, *B* seizes his hand and forces him to make stabbing movements which wound *C*. Again, *A* is suffering from St. Vitus Dance and in his tremors knocks over and breaks a vase. Lastly, *A* is driving a car when he is stung by a swarm of bees or hit by a stone; he swerves and runs over someone. The unconscious cases are perhaps more plausible. They include the following. *A*, driving a car, has a stroke without prior warning; he collapses and makes spasmodic movements at the wheel and runs down a pedestrian. Or, *A*, a woman asleep at night, starts to sleep-walk, takes an axe and wounds someone with it. Lastly, *A* unconscious, not in normal sleep but in a state of automatism, enters someone's house at night.

In these cases, conscious and unconscious, there is obviously something far more fundamentally wrong than where Smith shot and wounded Jones

owing to some mistaken belief that the gun was not loaded or that Jones would not be in the line of fire. But it is not at all clear at first sight what common features these cases are supposed to share. Why are they all treated by legal theorists as examples of the same fundamental defect — as cases where there is no 'voluntary conduct' or 'act of will'? What feature of normal conduct is missing here?

Now, in the effort to give some general answer to this question, the law books have resorted to a general theory of human action which can be traced back to writers of the early eighteenth century. This theory gives a very simple account indeed of what human action really is. Strictly and scientifically speaking, the theory says, action is, on its physical side, just a muscular contraction and nothing more, and the ordinary verbs of action in which we refer to actions, such as hitting, wounding or killing, are inaccurate and misleading. Why? Because they misrepresent as single actions what are in fact combinations of our muscular movements and their consequences, though only the muscular movements can properly be said to be 'ours' and things which we do. The theory admits that an action is not *just* a physical event, and claims that it is distinguished from a mere movement of our muscles by the fact that it has a special psychological cause, namely a desire for those movements, or, as the legal theorists sometimes term it, a 'volition'. This, according to the theory, is all that is meant by 'the will', and this is the crucial link between mind and body required even for strict responsibility. The volition is the element lacking in all our varied list of cases. So the problem

is solved by the very simple doctrine that in all normal action the muscular contractions are caused by a desire for them, whereas in our fundamentally defective cases they are not. Of course the theory concedes that in a full-blown action, according to ordinary speech, for example, that of hitting someone, besides the initiating muscular movements and the volition or desire for them, there will also be the foresight of consequences which is required in most cases for responsibility. But these further psychological elements are to be distinguished from the volition or will the presence of which is indispensable if there is to be an act at all.

The part of this doctrine which is interesting here is the contention that the natural description of our actions in terms of ordinary verbs like 'kill' or 'wound' is inaccurate, and all that we *really* do is to contract our muscles though usually with consequences which we also foresee. This is, I think, a profoundly mistaken view and my argument against it is parallel to that of the previous contributors. Thus, it is certainly sometimes true to say of someone that what he did was to contract his muscles. This would be so if, for example, he was under some form of physical training and was told by the instructor to contract the muscles of his arm. We might say, in this case, both that he contracted his muscles and desired to do so. We might say the same in cases where someone is baffled in some ordinary action like turning a door handle, if he adopts a special grip and exerts special force to get the thing to turn. But these are special cases distinguished by the circumstances and the unusual concentration on muscular effort. In

the ordinary case it is surely false to say that we launch our muscles into action by desiring that they contract. Indeed, in most cases we do not know what muscles we would have to contract in order to do a given action. If we are told to catch a ball or write down the letter 'q', few of us could say beforehand what muscular contractions we must make. Of course we might work the answer out either by imagining ourselves doing the actions and then thinking it over, or by doing them and then observing what muscular movements occur. But this is a sophisticated business and our normal awareness of our actions does not include awareness of the muscular movements involved. *A fortiori* it does not include any desire for them. Often, indeed, we may deliberate: Shall I hit him or not? But our deliberation is not conducted in terms of muscular movements. If any imagery accompanies our deliberations, what we see in the mind's eye is perhaps a picture of the *fait accompli* or some desired result.

It seems, therefore, that the attempt to characterize our fundamentally defective cases by saying that what is missing there is a desire, for the muscular movements must fail; for though it is quite true that no such desire is present in these cases, it is not present in normal action either. This cannot, therefore, be the dividing line between normal action and these grossly abnormal involuntary movements.

Must we therefore give up the conviction shared, I think, both by lawyers and by laymen that our cases, conscious and unconscious alike, do exemplify some single fundamental defect different from ordinary mistakes and accidents? I do not myself think we

should, for the intuitive feeling that there really is something which unites these varied cases seems to me to be sound. I cannot do more here than characterize in outline what I believe to be this common element; but I am sure that if we are to make progress we must start by abandoning the idea that the ordinary terminology in which we speak and think about our actions is in principle inaccurate and should be replaced by a terminology in which only our muscular movements are ranked as actions and all else are consequences. For the truth is, I think, just the opposite: if, but only if, we preserve the ordinary terminology which reflects the way in which our actions appear to us, we can then characterize this range of cases where the legal theorist says the element of will is lacking as cases where muscular movements occur which form no part of any action which the agent believes himself to be doing. This holds good, I think, both for the conscious and for the unconscious cases. For both where a man, startled or stung by a bee, jumps and involuntarily presses the trigger of a gun, or knocks something over in the tremors of St. Vitus Dance, and in the cases of epilepsy, sleep-walking or automatism, muscular movements occur which are uncontrolled, or wild, or not 'governed by the will', in the sense that they form no part of any plan of action on which the agent is consciously engaged. This, I suggest, is what we really mean when we say that the bodily movements are involuntary, or not an exercise of the will. The will as the controlling or governing agency is not to be conceived as a desire for muscular movements which causes them to occur, but as the mind of a

man consciously bent on doing something. This control may be lacking, either because he is not conscious at all, or because, though he is conscious, muscular movements occur which do not form part of any action which he believes himself to be doing. It is, I think, a merit of this way of characterizing involuntary movements that it preserves the distinction between such fundamental defects in conduct and those less fundamental aberrations from the norm exemplified in the case where Smith unintentionally wounded Jones, owing to the mistaken belief that his gun was unloaded. For in that case, though his wounding Jones was unintentional, his muscular movements were not uncontrolled or involuntary. For they formed part of some action which Smith believed he was doing, even if it was only the action of pressing the trigger of an unloaded gun. So the contrast between that case and one where Smith involuntarily presses the trigger when startled by a loud explosion is faithfully preserved in this form of analysis.

If analysis along these lines is right, I think a general lesson is to be learned from it. It is that if we are to succeed in clarifying the baffling notion of the will and its control over the body, and profitably to discuss the will's freedom, we must take care to start from our concept of action as it actually is and not from any rectified vocabulary or theory which falsifies the way in which our actions appear to us. And this I take to be a sufficient justification (if any were needed) for the concentration by contemporary English philosophers on the exact analysis of the complex terminology of action which we use.

4

DETERMINISM

GEOFFREY J. WARNOCK. 'Determinism' has meant so many different things to different people that perhaps it will be better to avoid asking straight out what it is. We should only find ourselves with a long list of not very clear theses. A list of that sort would be an interesting place to start a discussion from, but compiling the list and explaining the items would leave us no time for discussion of it. So I suggest we ask instead: How do people become determinists? What motives or reasons draw them in this direction?

J. F. THOMSON. Isn't it often that determinism of some kind seems to be enjoined by a scientific way of looking at human beings? Suppose you are an experimental psychologist. Then you may very well say: 'For my purposes, a human being is just something with an input and an output. And I am interested in discovering what correlations hold between input and output. Now the output at a given moment doesn't depend simply on the input then or immediately preceding. This is a fact of everyday observation. So I suppose that it depends on, at most, the input then and the state of the organism; the latter depending on, at most, past inputs, past outputs, and perhaps innate factors. But this means that the organism's output — what you call the man's

behaviour, how he responds to a situation — is completely determined by (*a*) what happens to him and (*b*) how he is.'

P. F. STRAWSON. And this picture completely leaves out and ignores choice.

J. F. T. It leaves it out. It doesn't deny it.

P. F. S. But you, quite naturally, made your experimentalist say: 'For my purposes.' This is surely important. If he had said: 'A human being is just something with an input and an output, that would have been a statement, a thesis, but it would almost certainly turn out to be either blatantly false or quite trivially true, because true by definition. But your experimentalist, I think, isn't asserting a thesis or not one about his subject-matter. He is saying that it is convenient to start off with a certain kind of model. He wants, for his own professional purposes, to regard human beings and their behaviour in a certain light. He certainly needn't say that this is the only way of regarding them; and obviously it isn't.

J. F. T. Yes, I think that is quite right. What I described could be called an engineer's way of looking at human beings.

G. J. W. But now doesn't this mean that the kind of 'determinism' you hinted at isn't a philosophical thesis either?

J. F. T. Well, I agree that if it is one, then it's very difficult to see just what thesis it is or what would count as a refutation of it. Our imaginary psychologist says that what the organism does at any given moment is uniquely determined by the situation it is in — or, perhaps, by how it 'sees' that situation, and also by the state of the organism. It

seems that 'determined' here has something like its mathematical meaning, as when we say that the volume of a sphere is uniquely determined by its radius. The volume is a function of the radius — a function of just one variable. So we should ask, just what are the variables on which the behaviour is held to depend? Since these aren't in fact specified, there isn't really a determinate thesis to discuss. It isn't enough to say that how a man behaves depends on something; you have to be able to say on what it depends.

G. J. W. And if the psychologist were actually to list some measurable quantities and then say that behaviour is some function of these and these alone, what he says might well be wrong — I mean untrue in a quite straightforward way.

P. F. S. Exactly. Whereas if he doesn't, if what the relevant variables are is left completely open, they might be anything — for example, they might for all that has been said, include variables connected somehow with acts of choice.

J. F. T. So it now seems that this kind of determinism is either not a thesis at all, or is a thesis so vague and ill-defined as not to be really discussable. Perhaps we should say that it is the form of a thesis.

P. F. S. Yes, but all the same you were right in implying that it must have seemed much more than this to many people. And we shan't have answered Warnock's original questions until we have seen why.

G. J. W. Yes, there is still the question why some form of determinism is so often taken for granted.

J. F. T. Well, I suppose that we are inclined to

extrapolate from particular examples of success in explaining and predicting human behaviour.

G. J. W. How do you mean? I might explain your behaviour to a friend in a quite satisfactory way simply by telling him what your own acknowledged reasons were for acting in the way you did. Or again, I might succeed in telling him correctly what your future course of action was going to be, simply by quoting your own statement of your intentions. These would be, wouldn't they, 'particular examples of success in explanation and prediction of human behaviour'. But it's hard to see how any amount of 'extrapolation' from such examples as these could lead one to determinism.

P. F. S. Would you really say that you were 'predicting' Thomson's behaviour when you merely reproduced his own statement of his intentions?

G. J. W. I don't see why not. I shouldn't merely be passing his statement on. I should be committing myself to the belief that he would actually do what he said he would do. In any case, even if the word 'prediction' isn't quite right for this kind of case, the word 'explanation' seems perfectly satisfactory for the other kind.

J. F. T. But it wasn't either of these kinds of case that I had in mind. It seems, on the face of it, that there are two quite different and contrasting ways in which human actions can be explained; and similarly two quite different and contrasting ways in which we can come to hold beliefs about what men's future courses of action are going to be. One of these ways in each case essentially involves going by a man's acknowledged reasons or disclosed

intentions. Insincerity and unforeseen obstacles apart, these are simply accepted as explanations of what was done or as statements of what will be done. The other way of coming to hold such beliefs is sharply contrasted with this. Acknowledged reasons and declared intentions either play no part at all, or play quite a different part. They are not accepted as the answers to our questions. At the most, they are treated as additional bits of the evidence in the light of which we frame our own answers. It was explanations of this kind which I had in mind when I said what I did say about extrapolation from successes.

P. F. S. If I understand you, you're suggesting two things: first, that the thesis of determinism is simply that the whole field of human action could be covered by explanation and predictions of your second kind; and, second, that particular examples of success in this kind of explanation and prediction incline us, or some of us, to think that the thesis may be true.

J. F. T. Well, for the moment, yes.

G. J. W. I'm afraid I'm not altogether clear. Is the suggestion that, if the second kind of explanation or prediction is available, the first somehow lapses or doesn't count, or embodies an illusion? Are the two kinds of explanation opposed? If you can explain an action of mine in the second way, does this mean that my giving my reason for it is no explanation at all? If you can predict an action of mine in the second way, does this mean that I am under an illusion in thinking that I decided to do it for certain reasons, that I was carrying out such-and-such an intention in doing it?

J. F. T. Well, it certainly seems that sometimes these consequences follow. We dismiss some reasons as rationalizations. But I don't think that the determinist — my determinist — has to claim that these consequences necessarily, or even usually, follow. Won't his point simply be that, on the hypothesis in question, we should be able adequately to explain and predict human actions without, as it were, taking anything on trust from the agent?

G. J. W. If this is all his point is, it doesn't seem to me to be enough to make him a determinist. When I rely on a man's announced intentions, I don't have to be taking anything on trust from him. I may simply know that he is not a man to misrepresent his intentions, or to change his mind unless his knowledge of the circumstances changes. Quite apart from announced intentions, I might very well be able to explain a man's actions and to predict, within rough limits, what he will do in important situations, without taking anything on trust from him, simply because I know him very well. Knowing him, I know what sort of reasons weigh with him, what sort of decisions he may be expected to make. This doesn't mean that the reasons don't really weigh with him, that he doesn't really make the decisions. And since it doesn't, I see nothing determinist at all in this possibility.

P. F. S. But I don't think this possibility can be quite what Thomson had in mind, though I agree it seems to be covered by the concessive form of words he adopted. When you spoke of knowing a man well, you were thinking, and indeed you spoke, of knowing what sort of reasons weighed with him,

what sort of decisions he might be expected to make. This is what we mean, or part of what we mean, by knowing a man's character. But suppose we had a different and deeper kind of knowledge. Suppose we were able to give a complete explanation, in terms of factors that didn't themselves include a man's reasons or decisions, of the fact that the reasons that weighed with a man did weigh with him. Then, in a sense, we could still admit that a man's action was explained by the reasons he gave for it; that these were his reasons. And this was the point of Thomson's concessions. But the fact that those reasons weighed with him would, by this hypothesis, be completely explained by factors of a quite different kind. And if an explanation of this second kind were available, the availability of the explanation of the first kind might seem to be a matter of smaller significance than it currently does. For we should have, as it were, two stages of explanation. There would be the question: 'Why did he act as he did?' answered by mentioning, among other things, the reasons that weighed with him. Then there would be the question: 'Why did those reasons weigh with him?' answered by mentioning only factors which were not reasons for him at all. If now we put the first question again — 'Why did he act as he did?' — we might well feel that we could give a complete answer to it simply by mentioning the factors referred to in the answer to the second question, together with a description of the stimuli the man was exposed to. The sense of the word 'Why?' will have changed, of course, just because, in the answer to it, we no longer mention the man's reasons. This is not to say that

he didn't have those reasons. Only now it will not seem so important that he had them.

G. J. W. That sounds more like determinism to me. But it also sounds enormously hypothetical. Is there any reason at all to suppose that one kind of explanation could be undermined in this way by the other?

P. F. S. Well, if you mean 'Could one kind of explanation ever be undermined by the other?' the answer surely is 'Yes'. Suppose that someone is, habitually, painfully shrinking and hesitant about entering a room full of strangers. We ask why. Well, his reason, the reason he gives, is just that to go into a room full of strangers frightens him; the situation just strikes him as frightening. But of course we want to know why he is frightened. Now the answer to this question, let's assume, is some fact about his early upbringing — some fact of which he may be completely unaware, so that it would not figure in the explanation given by him. Now, surely in this case, the fact about his early upbringing would quite naturally and properly be taken as the explanation of his odd behaviour; the explanation he gives, that he finds the situation frightening, would be regarded as having, by comparison, no explanatory force.

J. F. T. Yes, but I suppose the question Warnock had in mind was not whether this displacing of one type of explanation by the other ever occurs, but whether there is the slightest reason to suppose that it might occur universally, in absolutely every case.

G. J. W. Yes, that was what seemed to me enormously doubtful. And I think I can explain why.

what sort of decisions he might be expected to make. This is what we mean, or part of what we mean, by knowing a man's character. But suppose we had a different and deeper kind of knowledge. Suppose we were able to give a complete explanation, in terms of factors that didn't themselves include a man's reasons or decisions, of the fact that the reasons that weighed with a man did weigh with him. Then, in a sense, we could still admit that a man's action was explained by the reasons he gave for it; that these were his reasons. And this was the point of Thomson's concessions. But the fact that those reasons weighed with him would, by this hypothesis, be completely explained by factors of a quite different kind. And if an explanation of this second kind were available, the availability of the explanation of the first kind might seem to be a matter of smaller significance than it currently does. For we should have, as it were, two stages of explanation. There would be the question: 'Why did he act as he did?' answered by mentioning, among other things, the reasons that weighed with him. Then there would be the question: 'Why did those reasons weigh with him?' answered by mentioning only factors which were not reasons for him at all. If now we put the first question again — 'Why did he act as he did?' — we might well feel that we could give a complete answer to it simply by mentioning the factors referred to in the answer to the second question, together with a description of the stimuli the man was exposed to. The sense of the word 'Why?' will have changed, of course, just because, in the answer to it, we no longer mention the man's reasons. This is not to say that

he didn't have those reasons. Only now it will not seem so important that he had them.

G. J. W. That sounds more like determinism to me. But it also sounds enormously hypothetical. Is there any reason at all to suppose that one kind of explanation could be undermined in this way by the other?

P. F. S. Well, if you mean 'Could one kind of explanation ever be undermined by the other?' the answer surely is 'Yes'. Suppose that someone is, habitually, painfully shrinking and hesitant about entering a room full of strangers. We ask why. Well, his reason, the reason he gives, is just that to go into a room full of strangers frightens him; the situation just strikes him as frightening. But of course we want to know why he is frightened. Now the answer to this question, let's assume, is some fact about his early upbringing — some fact of which he may be completely unaware, so that it would not figure in the explanation given by him. Now, surely in this case, the fact about his early upbringing would quite naturally and properly be taken as the explanation of his odd behaviour; the explanation he gives, that he finds the situation frightening, would be regarded as having, by comparison, no explanatory force.

J. F. T. Yes, but I suppose the question Warnock had in mind was not whether this displacing of one type of explanation by the other ever occurs, but whether there is the slightest reason to suppose that it might occur universally, in absolutely every case.

G. J. W. Yes, that was what seemed to me enormously doubtful. And I think I can explain why.

You said earlier, you remember, that probably what inclines some people to accept a general determinist thesis is a simple propensity to generalize from examples of success in explaining and predicting human behaviour. But now, if we see what kind of explanation and prediction the determinist needs, doesn't there seem to be an objection in principle to generalizing from the examples of success? You see, the sort of behaviour Strawson mentioned in his example — excessive hesitancy about entering a crowded room — is abnormal behaviour; and this is just why the example suited his book. The reason the man gives — that the situation frightens him — is quite easily undermined as the explanation of his behaviour just because normal people are not frightened in this situation; so of course we have to look for a further explanation of his abnormality. Now I strongly suspect that our successes — successes of the kind that the determinist needs — are all achieved in this field of abnormal behaviour — typically psychotic and neurotic behaviour, and some kinds of criminal and generally anti-social behaviour. But if so, to generalize on the basis of these successes would be to propound the obviously absurd thesis that all human behaviour is abnormal.

J. F. T. Neat, but unconvincing.

G. J. W. Well, what's wrong with it?

J. F. T. Well, what worries me about it is really, I think, a general point about explanation. It's broadly true, isn't it, for fairly obvious reasons, that what we tend to have an interest in explaining is abnormality, deviation from the normally expected course of events. When things go as we expect them

to, we don't raise any questions — it's when something odd happens that we look round for an explanation. But of course that doesn't mean that the normal course of events can't be explained. When aeroplanes crash, we want to know why, we look for an explanation; when they don't crash no questions are asked. But of course there is a reason why aeroplanes usually stay up in the air and come down safely — that they do this depends, furthermore, on factors of exactly the same kind as those we invoke to explain why they sometimes crash. So the thesis you nearly presented as obviously ridiculous could be put in a quite sensible way — not as the thesis that all human behaviour is abnormal, but as the thesis that all human behaviour could in principle be explained in the same way as some kinds of abnormal behaviour can in fact be explained already. And that makes perfectly good sense.

P. F. S. But wait a moment. In making this general point about explanation, I think you're ignoring the special difficulty about human behaviour (as I think Warnock did too). People, unlike aeroplanes, often have reason for what they do, and often can say what their reasons are. Well, wasn't the significant feature of the example I produced, not just that this man's neurotic behaviour was abnormal (as Warnock said), but that the reason he gave for it was plainly inadequate? He is frightened, as he tells us — but by something which just isn't really frightening. It's not just that normal people aren't frightened of going into rooms full of people, but that there is no reason to be frightened. And that is why the reason he gives 'It frightens me' is regarded

as having no, or insufficient, explanatory force, and is accordingly displaced by explaining how he came to be afflicted with baseless fright in this situation.

G. J. W. All right, but I don't retract my point; I amend it. The position seems to be that the reasons people give for what they do are sometimes undermined by explanations of the other kind — which explain, in fact, why they are such as to be influenced by the reasons they give — but this second kind of explanation displaces or undermines the first kind in those cases in which (and precisely because) the reasons these people give are not good or adequate reasons for behaving as they do. But if so, to erect a general determinist thesis by generalizing on the basis of these cases still looks absurd: because it seems to imply the obvious falsehood that the reasons people give for what they do are never good reasons, and never have explanatory force.

J. F. T. But no, surely this isn't what the determinist is committed to saying. (I think I can now amend the point I made before.) It's admitted, or anyway it seems to be, that sometimes, when the reasons people give for what they do are very bad or inadequate reasons, we can further explain how they came to be influenced by such reasons — and in doing this we shall normally mention factors of a quite different kind from any they do give or could give, as their reasons; for instance, facts about their early upbringing, or perhaps about their physiological make-up. Now surely the determinist is not committed to saying that the reasons people give for what they do are never good reasons; but only that whether the reasons they give are good or bad, there

will always be a further explanation of the fact that those reasons weighed with them. And is there anything patently absurd in that suggestion?

P. F. S. No, I don't think it's patently absurd. But isn't it in some danger of turning out to be rather trivial? As the thesis stands, it contains too many blank cheques, so to speak, for one to be sure what it amounts to. We have to ask, surely, what kind of 'further explanation' is envisaged here. For example, if someone has, and avows, a perfectly good reason for something that he does, it is, I suppose, some sort of further explanation of his behaviour to say that he is quite an intelligent person, whose upbringing and background and present circumstances are perfectly normal and satisfactory, so that he is, in short, a rational man — the sort of man who tends to adopt good reasons as his reasons for doing things. But then two questions arise: first, it seems to me highly doubtful whether a 'further explanation' of this kind lends any support to any thesis that could aptly be called determinism — the points that add up to his being a rational man may explain his behaviour, but why should we say that they determine it? And, secondly, surely a 'further explanation' to the general effect that someone is a rational man has no tendency to undermine the explanation he himself gives of his own behaviour. Just the contrary, in fact; if the reasons he gives are good reasons, the fact that he is a rational man strengthens the view that the good reasons he gives do really explain his behaviour. So — either the 'further explanations' which the determinist claims could always be given do not (always) undermine explanations

of the kind the agent himself gives in giving his reasons, or the determinist must have in mind some special sort of further explanation, something different from the kind of thing I mentioned just now.

J. F. T. Well, he won't, I think, have in mind the kind of 'further explanation' you've just mentioned. If 'being rational' comes in, it will come in because he is someone who sets himself certain ideals, prides himself on — or even makes a thing of — rationality, and then this latter fact — supposing it to be one — will be put down to parental influence or something like that. But before we ask what kind of further explanation the determinist does envisage as being available in all cases, I think we ought to retrace our steps a little and ask what we mean by one kind of reason undermining another. I think you, Warnock, had in mind that the existence of the 'outside' kind of explanation (call it a B-explanation) would typically undermine the agent's own explanation (call this an A-explanation) in cases where the reasons given by the agent were not good ones. Hence you said that determinism as a general thesis seemed to imply 'the obvious falsehood that the reasons people give for what they do are never good reasons'. But surely there is quite a different way in which the existence of a B-explanation might undermine or discredit an A-explanation; it might tend to show that the reasons given by the agent, whether adequate or inadequate, weren't really his reasons, or weren't all of them.

P. F. S. I think what you're now calling attention to isn't, or isn't only, an obscurity in the notion of undermining; it's a general difficulty

about A-explanations. If we ask someone why he did something, he may reply by 'giving his reasons' — or by 'giving the reasons'. In the former case, what is in question is, roughly, whether what he did was a reasonable thing to do; he will aim to show that it was, and to do this by telling us of good reasons for doing it. And his standpoint here is that of an agent who is, or wants to be thought, reasonable and responsible. But in the second case, he is not concerned with the reasonableness of the action, but only with explaining it.

G. J. W. Is this distinction as clear as you suggest? Take 'I hit her because she was becoming hysterical'. That she was hysterical is here put forward as showing that an action was reasonable and justified. But it also explains it in a quite straightforward way.

P. F. S. But I think all that matters for our purposes is that we can distinguish two ways in which such a statement might be attacked. It might be attacked, first by denying that she was becoming hysterical, or by denying that this is the right way of handling someone who is becoming hysterical. Or, quite differently, it might be met with the suggestion that it doesn't correctly or doesn't completely state the agent's motive. For example, it might be suggested that although she was becoming hysterical and although this was the best way of handling the situation, the man in question hit her partly because he wanted to or because it gave him pleasure.

J. F. T. Exactly. And the reason I wanted to draw this distinction is that it enables us to see that determinism needn't have the absurd consequence

that Warnock mentioned, that no reasons are really good reasons. The consequence that would presumably remain — that many and perhaps all of the assessments we make of our own motives, and our statements about how we came to do the things we did, are at best incomplete and partial — this consequence doesn't seem absurd, at least to me.

P. F. S. But I still don't see just how every such statement could turn out to be 'incomplete'. And this brings us back to the question I asked: what sort of further explanations does the determinist envisage being available in every case?

G. J. W. Your question is a pressing one, I think. It's now generally recognized that behind, or besides, acknowledged reasons there sometimes lie unacknowledged reasons — generally of a less reputable kind — of which the agent may not be wholly conscious or of which he may be wholly unconscious. But, first, it would seem difficult to generalize this thesis without trouble; and, second, and more important, it is hard to see that it offers even limited support to determinism. When a man is brought to acknowledge an unacknowledged reason as his, he certainly sees his action in a different light; but not as an action for which he can disclaim responsibility.

J. F. T. Perhaps the would-be determinist might look altogether elsewhere for his further explanations. What about this? All human action is ultimately a matter of physical movement, of movement of parts of the body, including, of course, the organs of speech. All such movements may be completely explained as the consequences of events within the

body, of the kind studied in physiology and neurophysiology. And all these events may themselves be completely explained, in terms of physical laws, as the result of physical stimuli acting upon a body of a certain physical constitution. Is this deterministic enough for you?

G. J. W. You would agree that you describe at best a programme rather than an achievement?

J. F. T. Certainly. But not an absurd one.

P. F. S. I think you must agree too, that it's a different sort of programme from the one you initially mentioned. Initially you spoke of finding explanations of why certain reasons weighed with a man, in the light of which we might be inclined to think of his action as wholly explained by factors which were not reasons for him. The reasons were to be first explained, then by-passed. But the new programme doesn't even mention reasons. It just envisages physical explanations of physical movements.

J. F. T. Yes. I said, I think, that the would-be determinist might look altogether elsewhere for his further explanations.

P. F. S. I think the new programme is rather a lot to take on at this stage. But my first reactions are these. First, and concessively, it seems to me pure folly for any philosopher to declare, about any scientific aim that can intelligibly be stated, that that aim is in principle unattainable for philosophical reasons. But I think that what you said wasn't simply the statement of an intelligible scientific programme. It began with the purely philosophical remark that all human action is ultimately a matter of movement of parts of the body. It is this which

makes me suspicious. I agree that we might, for some purposes, describe a particular bit of human behaviour in terms of 'moving parts'. But I don't agree that such descriptions would be adequate to the concept of human action. So long as we are the sort of creatures we are, the possible availability of descriptions of particular bits of human behaviour in these terms wouldn't dispense us from the need to see and interpret human actions in terms of a quite different set of concepts. The first sort of descriptions might indeed all find a place in, or under, physical explanatory laws of the kind you envisage. But the second and indispensable sort could find no such place. They belong to a different kind of vocabulary and call for a place in a wholly different dimension of explanation.

J. F. T. But this sounds like a mere evasion. Wouldn't they just be the same events under different names? If I completely explain an event under one name, calling it a different name doesn't mean that I haven't completely explained it.

G. J. W. Isn't Strawson's point precisely that it wouldn't be simply a matter of different names for, or descriptions of, the same event? The two names, as you call them, belong to wholly different systems of classification. Unless those systems of classification exhibited a kind of correlation which we can be quite sure they won't exhibit, then however much neuro-physiology advances, no interesting thesis of determinism can be based on that advance. There may be as complete a system as you like of physical laws of movements of bodily parts; but there will still be no physical laws of human action.

J. F. T. But I'm afraid this still sounds to me like a mere evasion. You seem to me to be saying that the thesis that all movements of bodily parts might be brought under a system of physical laws is, in a sense, not interesting, for the reason that the system (if we can call it such) within which we employ the concepts of human action is, and would remain, a quite different system. Well, that can scarcely be denied. But aren't you also assuming that these systems are not just different, but also independent?

P. F. S. How do you mean?

J. F. T. Well, you seem to me to be suggesting that we now have a way of talking about, and a way of seeing, human behaviour (bringing in such ideas as those of reasons, intentions, choices, decisions and responsibility) — and that it would just make no difference to this whether or not it were also possible to explain all bodily movements completely in physical terms.

P. F. S. Well, what difference would it make?

J. F. T. I'm sure you'd really agree, anyway, that it isn't just obvious that it wouldn't make any. What about this, for instance (though I'm talking pretty vaguely here) — mightn't one say that our current way of thinking and talking about human action tacitly makes a claim (which, by most people, is tacitly conceded) to being indispensable to the explanation of what goes on in the world, so far at least as human beings are concerned in what goes on? That is, isn't it perhaps taken for granted that where human beings are involved, what goes on can neither be predicted, nor explained, without em-

ploying the notions of reasoning, choosing, deciding and the rest. But your thesis denies this, at any rate; and that's quite enough to make it far from uninteresting.

P. F. S. I really must disagree with you about this. You speak of 'what goes on, where human beings are concerned'. But what does go on? Certainly, bodily parts move. Equally certainly, people do things. The blanket expression, 'what goes on', conceals the fact that there are no effective correlations between the two vocabularies for talking about what goes on, the vocabulary of human action and the vocabulary of physical science. If, for example, 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 girl friend were an instance of another such class (and vice versa), then the reign of law in the field of 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 no question of any such correlations ever being established. And it's no good saying that every description of a type of human action must in fact be correlated with some disjunction or other of descriptions in purely physical terms. This would be a pointless piece of logic, which could never be the basis of a scientific theory of human action, and could never yield deterministic explanations of human action.

J. F. T. Well, I think I must disagree with you in turn. First, if we assume, as it were, unlimited

success in physical explanation of physical movements, then it's by no means clear to me that people in general would find your point about absence of correlations an adequate defence against determinism. To many this defence will seem rather like surrender. And, second, why are you so sure that 'effective correlations', as you call them, can't be established? This sounds like a piece of dogmatism, on a matter which should be left to scientific investigation.

P. F. S. On the first point — I can't help it if people are confused. And on the second — I don't think it is a matter for scientific investigation, but for philosophical argument.

G. J. W. I'm afraid it's too late to pursue this disagreement any further. I think perhaps we ought to try to summarize what's emerged so far. We've referred to more than one type of thesis as deterministic, or possibly deterministic, and I think we ought to end by listing and distinguishing them. The thesis we've just been discussing might be labelled 'physical determinism'. This was the thesis that human behaviour, at least in so far as it consists in movements of bodily parts, might be brought under a system of purely physical laws. This was agreed to be a programme, not an achievement. But there seems to be room for disagreement as to whether even the fulfilment of the programme would establish what might be called a physical determinism of human action. Perhaps, for this, some other conditions would need to be fulfilled. And there seems to be room for disagreement about whether they could be fulfilled.

J. F. T. What we were discussing earlier, of course, was something quite different from this — it could be called psychological determinism.

G. J. W. Yes — but it's important to notice that we actually considered two possibilities, both broadly 'psychological' I suppose, but different. The first was the thesis that explanations of human behaviour in terms of the agent's own reasons could always be supplemented by explanations, in quite different terms, of the fact that those reasons weighed with him; and what we chiefly discussed here was whether, and, if so, in what sense, this kind of supplementary explanation could properly be regarded as undermining, by-passing, or somehow displacing the other kind. Then Thomson mentioned the rather different thesis that the reasons an agent gives for his action might be held to be always incomplete — that he always has further, unacknowledged, perhaps unconscious reasons (or motives). What I said about this, you remember, and I still think rightly, is that, although no doubt it is a somehow worrying thesis, it is very doubtful whether it should be called a species of determinism; and also very doubtful whether it could plausibly, or even reasonably, be put forward as true of all human behaviour.

P. F. S. But that last remark, of course, applies to any kind of determinism which is really a thesis, and not just a programme. Perhaps the clearest point of all is that no such thesis worth taking seriously is known to be true.

J. F. T. Or false.

5

ACTIONS AND EVENTS

WHEN we discuss the question whether what someone did (perhaps ourselves) was the right thing to have done or not, was prudent or ill-judged, sensible or silly, a question that is always liable to come up is the question what the person concerned *could have* done, instead of what he actually did. And the fact that this question may and often does come up is, of course, absolutely central to the way in which we think and talk about human behaviour, including our own. What people do is often, perhaps more often than not, what they choose to do from among the things they *can* do; or even if they don't actually choose, don't really pause to consider alternatives, we think that often they do have a choice, alternatives are open. It is, in fact, only because we think this is so that we hold people, ourselves included, responsible for what they do, and liable to be blamed or commended, punished or rewarded, for their behaviour. This fact (if it is a fact) has been picked out particularly by the Existentialists as *the* characteristic fact about human beings; unlike objects, they don't *just* exist, *just* undergo some sequence of happenings, but can and must *choose what* to be, or what to do, from a range of available possibilities. This fact may not strike us all, as it seems to strike Existentialists, as 'nauseating';

still its central place in our thought and our behaviour is indisputable.

If so, it is clearly a matter of some importance that this central fact should really *be* a fact, that it should actually be *true* that, sometimes at least, people can do things other than the thing they actually do, or could have done something other than the thing they actually did. Is there any reason to be doubtful about this? Well, of course, some philosophers and other theorists have argued that there is: the notion that a person could have done what he did not do, or more generally that anything could have *happened* which in fact did not, is inconsistent, so they maintain, with 'determinism'; and determinism is supposed to be a thesis which at least may be true, and (some have said) is true. Now the question I want to discuss is not the question whether determinism is true, but the old question whether it — or some version of it — is or is not *incompatible* with what we all ordinarily believe about human behaviour — with the idea that, often, people *could* have done what they did not do. The conclusion that, in many ways, one would like to come to (and which in fact has had sponsors here and there since at least the time of Hobbes) is that there is *no* incompatibility here; for it's only if our ordinary beliefs are contradicted by determinism that we need really *worry* about whether or not determinism is true; if our beliefs can stand even if determinism is true, then the question of its truth or falsehood can be regarded as comfortably academic, or perhaps even abandoned as just not interesting at all, as not worth discussing.

One might set out our question then in the

following way: To say that someone could have done something which in fact he did not do, is to say that there was *no bar* to his doing that other thing, nothing to *prevent* him from, or to preclude his, doing it. Put this way, our main question looks as if it should be simple. If determinism is incompatible with our ordinary beliefs, then it must somehow maintain that there is *always* a bar, whereas we ordinarily suppose that sometimes there isn't; but if determinism does *not* assert or imply this, then our ordinary beliefs about human action are in no peril, from this quarter at any rate. And surely we should be able to decide between these alternatives. Well, perhaps; but the decision is far from being an easy one to make.

Let's say — over-simplifying a good deal for the sake of brevity — that barriers to action can be classed under two headings: those which consist in facts about the person concerned and those which consist in facts about his situation. Suppose, for instance, that a Russian guest at dinner in my College inadvertently sprinkles a lot of salt on his pudding, mistaking it for sugar, and so renders his pudding markedly unpalatable; and suppose that for some reason I am reproached for this *contretemps*, and am told that I ought to have prevented it. I reply 'But I *couldn't* have prevented it'; and I proceed to cite either (*a*) a fact about myself, that I can't speak Russian, or (*b*) a fact about the situation, that the unfortunate guest was too far away to hear anything I might say; or perhaps both. I then add that my colleague Jenkins *could* have prevented it; for he both had the ability to make himself understood in Russian,

and was actually sitting next to the unfortunate victim. So, whether or not one takes the view that he *should* have intervened, at least he could have; both his abilities and the situation rendered that course open to him.

(I should perhaps just mention that this distinction between abilities and situation should not be taken too seriously; 'abilities' here, for one thing, is a kind of short-hand for abilities, skills, capacities, powers, talents, rights, privileges, and no doubt other things still; and for another, it would often be a matter of indifference whether some fact, for instance about somebody's *official* powers, was regarded as a fact about *him* (his being an official), or about his situation (his being in an 'official position'). Still I think we can legitimately use this rough distinction as workable enough for the present purpose.)

Our views, then, about what somebody *could have* done can be said to turn on our beliefs about his abilities and his situation. These are not, of course, always *equally* problematic, even if the question strikes us as problematic at all. For sometimes the alternative action in question, about which we ask whether he could have done it, is one that is within the capacity of practically everyone to perform, so that the only serious question is whether it could have been performed in the circumstances; sometimes it will be plain, or at least seem so, that the thing was possible in the circumstances, the only question is, had the man concerned the necessary ability? And very often, of course — or, at least, so we suppose — *neither* of these questions is in any doubt at all. He

had the ability, the situation set no barriers in his way, so he could have done it; if he didn't it wasn't that he *couldn't* have done it. Perhaps he just decided not to, or to do it just never occurred to him, or something of that sort.

Now one way, obviously, in which we may come, and in fact have come, to change our views about what people can and can't do, could or couldn't have done, is by coming to change our views about their abilities or powers. Our ordinary common-sense, common-knowledge ideas on this subject have been (or at any rate should have been) modified by, for instance, psychological researches, which have shown that there are more cases than we might have thought in which people are genuinely not *capable* of behaving in certain ways, however much we and even they may wish that they were. Might a determinist, then, undermine our customary beliefs by contending that this trend, this progressive shift in the assessment of human capacities, might be continued to the limit — to the point at which we come to understand that no one *ever has the ability* to do anything except just what it is that he does?

But no: there is surely an absurdity in this suggestion. For an ability — at least as we have used the term so far — is an ability to do a certain *kind* of thing, for example, to speak Russian; and of course there will always be cases of particular things we don't do, which are yet things of the *kind* which we certainly do have the ability to do; perhaps we often do them, though not, as it happens, on this occasion. Furthermore, it isn't really at all plausible to suggest that advancing psychological knowledge *will* always

tend to restrict the scope of human abilities; may we not sometimes, in virtue of the discoveries of psychologists, come to be able to do things that previously we could *not* do? Thus, there seems no reason at all to suspect that, at least from the side of our *abilities*, anything whatever could tend to show that we are mistaken in thinking that sometimes people could have done what they did not do; that they *had the ability* to do something else will often be true beyond the faintest possibility of questioning; and if determinism denies *this*, then determinism is absurd on the face of it.

It appears, then, that if determinism is to be a serious threat to our beliefs about human behaviour, the threat will have to take the form of denying something that we ordinarily believe about the *circumstances*, or *situations*, in which people act. We are not always radically and completely wrong in our estimation of human abilities; but perhaps we *are* in some way radically wrong in our beliefs about our situation. So is there, in fact, anything in any deterministic thesis worth taking seriously, that, if true, would oblige us to admit this?

Now it is in some ways a curious fact that this simple-looking question is extraordinarily difficult to answer — though, of course, the long and tangled history of controversy in this area strongly suggests that this must be so. The trouble is that deterministic theses, even relatively clear ones, impinge upon our ordinary beliefs, if at all, in a bafflingly oblique way. You may remember that this difficulty came up towards the end of the foregoing discussion on determinism, and I want now to take up again a point

which was there in question, but not very thoroughly examined.

We were discussing at that point a thesis which we dubbed 'physical determinism' — the thesis, baldly stated, that human bodies are just like any other physical objects, in the respect that everything that happens to and in them can be brought under a deterministic system of physical laws — the physical state of the body at a given moment is the unique resultant of its state at preceding moments, perhaps modified by the intrusion of external physical stimuli. Now Mr. Strawson pointed out, as you may remember, that contrary to what has commonly been assumed, this thesis does *not* say that *human behaviour* is physically determined; it says, in fact, *nothing* about human behaviour, since it is expressed solely in terms of the states and movements of bits of matter, and no expression describing any piece of human behaviour says *merely* that certain bits of matter moved in a certain way. Furthermore, Mr. Strawson said, and I think he was right, that there is no prospect of correlating descriptions of human behaviour, whether uniquely or disjunctively, with specific statements about the states and movements of bits of matter, in such a way that one could *derive* a doctrine of determinism of human behaviour from the thesis that the physical organism is deterministic. Now I think he meant to imply that, for this reason, even so radical-looking a thesis as that of physical determinism could not shake the beliefs that we have about human behaviour; for it does not (the suggestion is) impinge upon them; we want to say that sometimes people could have *done* what they did not *do*, but the thesis

says simply nothing at all about what people do. It is merely a thesis about how matter moves. If so, then this, which is perhaps the clearest and most plausible version of determinism, seems (consolingly) *not* to go against what we wish to say about human action.

Now I am very, very far from being clear about this: but the fact is that I am not to be thus consoled. It at least seems to me that there is a complaint which many people would be tempted to make here, and I am by no means sure that the temptation ought to be resisted. The complaint is this. The point that specific statements about how matter moves are not, and, furthermore, are not translatable into, specific statements about what people do, is a correct point, and a substantial point. We have here, as one might put it, fundamentally different systems of concepts, between which no tight logical relations obtain. But still, even if for this reason we concede that the thesis of physical determinism neither states nor entails that human behaviour is determined, one may still feel that there are *some* beliefs about human behaviour which that thesis, if true, would require us to abandon.

Consider an example. Suppose that, in my rôle as physical determinist, I describe a certain process of matter in motion, which is in fact (though I don't use these coarse, everyday expressions) that of Smith's foot coming into contact with Jones's shin; and I claim that, given the antecedent physical set up, matter *had* to move thus, nothing else could have occurred. Now in saying this I do not tell you *what Smith did*, and *a fortiori* I don't tell you what he could

not but have *done*. There is indeed a sense in which I don't talk about Smith at all. He might have been kicking Jones on the shin, certainly; but, for all I have *said*, he might equally have been embarking on a dance in the course of which his foot knocked Jones's shin accidentally; or clumsily showing to Jones his new shoe-laces; he might even have been *doing* nothing, being the helpless subject of stimuli he was powerless to resist or control. And so on. But, even if the account of how matter moved thus does not entail any specific designation of what Smith did, there still are, surely, *some* descriptions of what he did with which it is incompatible. For instance, if matter moved in the way described, then, though it is not necessarily true that Smith kicked Jones on the shin, at least it *can't* be true that he kept both his feet on the floor. It appears — indeed, surely it is obvious — that the narratives of the physical determinist will be compatible only with those accounts of what people do which are themselves compatible with matters moving in the way narrated: and though there may well be several alternative such accounts, there will certainly be *some* that won't do. But this seems to mean that, if the thesis of physical determinism is true, then there *are some* things which, in a given situation, a person could *not* have done — those things, namely, which are ruled out by the narrative of how matter moved. And this does not seem to leave us anything like as much *latitude* in our views about what people could have done as we want to have and as we ordinarily believe that we do have in fact. Indeed I think the situation is really rather worse than this.

Remember that we are not now discussing the individual's general *abilities*. We are assuming it to be agreed that people very often have the general *ability* to do things which they do not do, and raising only the question whether, as things then were, an individual ever could *in fact* have done anything other than the thing he did. Now, if the thesis of physical determinism is true, we can answer 'yes' to this question only, it seems to me, in a very oblique, back-handed and indeed unnatural sense. Given the physical set-up, nothing could have occurred except what did occur; that is, matter could have moved *only* as in fact it did move. Now what the man in question *did* is certainly not thus determined, but the question left open here is surely only this: what, given that matter so moved, *might the person have been doing*? Certainly there will be a variety of possible answers to this question. But it seems clear that the range of possibilities will not be *the same* as that which we should normally consider as possible answers to the question *what he could have done*. In accepting the thesis of physical determinism we have in effect confined ourselves to those descriptions of actions *only* which are compatible with matters moving in a certain way; and in any case, it is surely evident that the question what someone, whose limbs moved in a certain manner, might have been doing, is, though open enough, quite different in principle from any ordinary sense of the question, what he *could have done*. We do not, if so, really make room for this latter question by showing that the thesis leaves room for the former one, and our denial that human behaviour can on these grounds be said

to be determined turns out to have undergone a rather subtle, but I think a most important, shift of sense.

Or *is* it important? Does it, after all, *matter* whether we accept or reject this so-called 'physical determinism'? The thesis that every movement of matter, including those bits of matter which constitute human bodies, is uniquely determined by physical laws and an antecedent physical state may be thought to be a perfectly *useless* thesis, since perhaps we could neither hope to show that it was true nor take any practical account of it even if it were true — in particular cases the relevant facts might always be, in practice, impossible to establish. Still, it *would* be, I think, a point of some importance if, as seems to me to be the case, our ordinary beliefs about human behaviour were inconsistent with it. For, if so, and if we should wish to retain those ordinary beliefs, we would seem thereby to be committed to a belief about the physical world — very roughly, the belief that it is not 'self-contained' — which has the exciting character of being quite widely regarded as absurdly superstitious. However, I am more than prepared to be persuaded that I am in some way quite mistaken on this topic, and that accordingly I must deny myself here the excitements of superstition.

6

FREEDOM AND KNOWLEDGE

D. F. PEARS. We are going to discuss the connection between knowledge and freedom of the will. When people ask how these two things are connected, they usually approach the question in a rather external way. For the agent is usually seen through the eyes of spectators who are credited with a great deal of knowledge about him. How, it is then asked, does their knowledge bear on his freedom? And the determinist's answer, which has received a good deal of attention in these essays and discussions, is that, on the whole, the spectators' knowledge bears unfavourably on the agent's freedom.

But the agent also has another point of view, which differs from the spectators' point of view. For the agent often knows that he will do something simply because he has decided to do it. And this special kind of knowledge, unlike the scientific knowledge of spectators, might have favourable consequences for the freedom of the will. For we might find a way of avoiding determinism if we approached the question from the point of view of the agent himself, as Professor Hampshire does in his recent book, *Thought and Action*. We might find that a person's freedom of action is not necessarily impaired by the possibility that spectators might predict what he will do. In fact, if he knew about the spectators' predic-

tion, his freedom of action might actually be increased; for if the spectators based their prediction on a known tendency of his, he might be put on his guard against this tendency, and so might be better able to resist it. Of course, the spectators might predict that he would not even try to resist the tendency. But perhaps he himself could not agree on scientific grounds that he would not try. For it is arguable that if he agreed that he would not try, this would really mean that he had decided not to try.

Hampshire, in his book, uses considerations like these to show that there is always an area in which a person cannot say, on scientific grounds, that he will act in a certain way: an area in which he can know what he will do only in the special way that comes from deciding. And we are going to examine his argument in order to see what bearing it has on the question whether the will really is free.

P. L. GARDINER. It might at first sight seem unplausible to suggest the agent must necessarily take a different view of himself and of the things he does from the view taken by an 'external' observer, that is, by someone different from himself. For — as you pointed out — a person may have his attention drawn by somebody else to certain of his tendencies or character-traits, and there seem to be no grounds for thinking that the person himself might not become aware of such tendencies as a result of his own self-observation. For example, he might recognize that he regularly tends to respond to situations of, say, an emotional character in a certain way — he constantly finds himself being drawn into personal relationships or entanglements of a particular type. And

this knowledge (which he may or may not be prepared to admit publicly) he presumably acquires in ways that are (to a large extent, at least) similar to the ways in which he becomes acquainted with the traits, tendencies, and so forth, of those around him. If he were to ask himself how he knows the things he does about his own disposition the most obvious answer that might occur to him would be that past experience showed them to be true. But if this is so, what — it may be enquired — is to prevent him from utilizing such previous experience as a means of, or guide to, predicting what he *will* do, or try to do, when situations similar to those he has encountered in the past, present themselves to him in the future? In other words, can there be any objection in principle to such a man's saying that, knowing himself as he does, he 'expects' or 'foresees' that he will (or would) act in such-and-such a way in a certain contingency? Aren't such predictions in fact quite commonly made? But I think that you, Hampshire, would nevertheless want to claim that this rather simple view is not free of difficulty.

S. N. HAMPSHIRE. I agree — and I make the point in my book — that all of us can observe our own behaviour, as we would observe the behaviour of another, noticing regularly repeated patterns; and that we can apply this knowledge in order to make predictions about our own future actions. But within what limits can we apply this inductive knowledge to ourselves, in predicting what we shall try to do? Suppose that I am not sure whether to undertake a difficult piece of work. I can perhaps compute, from experience, what my chances of success

are, if I do decide to try. But this leaves the question — Shall I try? — open for decision. Now, I might know, as a fact about myself, that I always vacillate and am uncertain when faced with decisions of this kind, and that, after a period of vacillation, I always in fact choose the easiest, safest course. Might I not then say 'I have not made up my mind yet: but I know from experience what I will do in the end'? I do not deny that a man may use these words: but the question is — What sense can he attach to them? Surely if he still has to decide what he is going to do, he must still be uncertain what he is going to do. This seems to me a logical (or conceptual) necessity. In my book I argued that a man cannot — 'cannot' in the logical sense — claim to know, on the basis of inductive evidence, what he will himself try to do *on a specific occasion*, and at the same time claim that he has not yet made up his mind, or decided. Inductive prediction of one's future own actions, I argued, is confined to the following three cases: (1) predictions of success or failure in some attempt; (2) prediction that one will, at some time unspecified, behave in a certain way; (3) prediction that one will act in a certain way, where the action is not entirely voluntary and performed of the agent's own free-will.

D. F. P. I'm not sure about this. You say that the agent cannot predict that he will make up his mind to do something on a particular occasion and try to do it. The case for saying that he cannot make this prediction is that it includes his own future decision; and so, presumably, includes his reasons for his future decision; and so, you would argue,

when he makes what looks like a prediction, he is really endorsing his future reasons, and therefore really deciding in advance. And a decision, taken in advance, is not a prediction.

But isn't there sometimes room for psychological self-prediction in such cases? For example, a person sometimes knows that a particular consideration will seem to be much stronger when the moment for action arrives than it now seems to be. And doesn't this show that he can predict that it will eventually lead him to make a decision, although he hasn't yet made the decision?

I think that you are assuming that the weight of rational considerations for a given person is constant through time, so that, if he predicts that a reason will seem to him to be conclusive at some later date, he is already endorsing it as conclusive. But isn't a person's reasoning subject to vicissitudes which he himself can sometimes predict?

S. N. H. In your example, the man *knows* that a particular consideration, or reason for acting in a particular way, will seem to him stronger at the moment of action than it does now: and he knows this from his observation of himself, and of his own processes of decision in the past. In what sense of 'stronger' does he know that it will seem to him 'stronger'? When he predicts this seeming-a-stronger-reason-to-him, is he predicting his state of mind at the moment of action? Certainly there is a sense of 'state of mind' in which he may predict on the basis of experience what his state of mind will be at the moment of action: that he will be tired, or that he will be in a disillusioned or excitable mood,

and therefore that this consideration would probably influence him unduly — unduly, that is, as he judges the matter now. But he is still uncertain what he will try to do on this occasion, since he is still considering the question, and is still undecided: he still has not predicted, or said to himself, or settled in his own mind, what he will in fact try to do: he has only predicted something about the *circumstances* in which the decision will be made at the moment of action, if it is not made before. Suppose he used the following form of words: 'I know that this consideration will outweigh all others, when the moment for action comes, because I shall be so tired and disillusioned: therefore I predict that I shall try to do so-and-so'. Should we not say: 'Have you then decided to let the events — in this case, psychological events — take their course?' This also is a decision.

D. F. P. I would like to define more precisely the nature of the example that I was putting forward as an exception to your thesis. You say that what your tired man eventually does, or at least attempts to do, is the outcome of his psychological state, and not of the reasons that he adduces. But this is ambiguous in a highly important way. It might mean that his reasoning had nothing to do with his ultimate attempt, since the immediate determinant of his ultimate attempt was his psychological state, which would have had the same outcome whatever reason he adduced, and perhaps even if no deliberation occurred at all. Alternatively, what you said might mean that his reasoning was an essential link in the chain of determinants, but that it was itself influenced, perhaps for the worse, by his psychological

state. Let us say that in the first case the reasoning and decision are inoperative, and that in the second case they are operative. And the example that I was putting forward as an exception to your thesis belongs to the second type, where the reasoning and decision are operative, and the self-prediction goes *through* them.

I would also like to make another distinction. You say that in this example, when the man predicts what his decision will be tomorrow, he is thereby deciding to let the psychological events take their course. Now, if he were making this decision, it would not be a decision about the same thing as his decision on the following day. For tomorrow he will decide the primary thing, what to do : and today, according to you, he is making a secondary decision not to try to make his deliberations take a different course tomorrow. And this distinction between primary and secondary decisions is an essential one, and you use it in your book when you apply your thesis to the question of freedom of the will.

IRIS MURDOCH. I don't see why a man who predicts how he will decide tomorrow need always be thought of as having *ipso facto* decided not to try to alter the course of events leading to that decision. There *are* cases where one could say that predicting puts one in the position of having decided — as when a man predicts that he'll be too tired to think clearly when the time of decision comes. His prediction may suggest that he is tolerating the state of affairs he envisages, and isn't going to try to alter the intervening course of events. But predictions needn't always be concealed decisions. A man

might predict that when it came to the point he might be influenced by pity and not insist on justice — and might feel it proper to be quite neutral about this. He might predict that he'll reason differently later, and he might simply note this possibility, without taking sides. And if he's not taken sides he has not decided.

Reason and feeling are not always as separate as Hampshire's picture suggests. It may sometimes be quite proper to envisage that one will 'feel differently when the time comes' without now deciding whether or not to try to make oneself feel differently.

Anyway, does Hampshire's thesis require that there *must* be a decision *whenever* there is a prediction? Would it not be enough to say that one always could, having predicted a later decision, make a move which would falsify the prediction? Put in this way the thesis begins to look rather more like a statement of fact, and rather less like a conceptual necessity. If one is told 'You always can interfere', one may start wondering to oneself, 'Well, can I really?'

I'm still not sure what the 'logically necessary' aspect of Hampshire's argument is supposed to be. Is it this, perhaps: that *if* a man pictures himself he (logically) *must* picture himself as able to interfere with whatever he predicts about his own voluntary action. This would leave open, first, the possibility of his picture being simply wrong. I'm not sure what Hampshire would say about this. Then, secondly, there is the possibility of his not picturing himself at all. If a man does not picture himself as an agent (doesn't know what he is doing or that he's doing it), I think Hampshire would regard him as not

sufficiently rational or self-aware to exemplify his proof of freedom. But what about the third possibility: the case where the man does reflectively picture himself, but pictures himself differently from the way Hampshire describes?

One is tempted to argue, and this might *look* like a logical argument to a truth about human nature, that beings convinced of a deterministic status would have a different and somehow non-human mode of consciousness. But it's not clear to me that this need be so. Not all men view 'will', 'choice', 'decision', exactly as we do. Modes of consciousness may differ widely, which is why it's difficult to construct a *logical* argument to a general truth about human nature.

S. N. H. I accept what you say about 'envisaging the situation'. I have said that it is complementary to deciding in the sense that to decide is either to decide to intervene in some way in the envisaged natural course of things, or to decide to let things take their natural course. The difficulty of drawing the line, in any particular case, between envisaging and deciding, is that envisaging may be of different degrees of explicitness and definiteness, and consequently so also may deciding. The only 'ought' that enters into my analysis is the requirement that we ought to be as clear and definite as possible in the envisaging and the deciding, and consequently that we ought to know which we are doing: it is a kind of evasion to blur the necessary contrast, to hide it.

Now, as Pears suggested, my thesis can be stated in terms of the distinction between primary and

secondary decisions. The agent, I am supposing, is wondering what primary decision to act he will make in the future. And I am assuming that this primary decision will be operative — *i.e.* that it will make a difference to his ultimate action. Now, if a man has made a decision, or formed an intention, he knows what he will do, or at least what he will try to do. For intention implies knowledge of what one is doing or of what one is going to do. So the man who *knows*, as a matter of psychological fact about his own dispositions, that he will be influenced by pity at the moment when he makes his primary decision, will be said to have allowed himself to be swayed by pity, if he does not now make a secondary decision to take (if he can) some action to counteract this effect: and so he may be said to have made a secondary decision to allow himself to be swayed by pity, if he in any way actively considered the issue. Otherwise how could he be certain what he will do?

His knowledge of the natural course of things — inductive knowledge — certainly includes knowledge of his own mental processes, including his decision-making processes. But in deciding upon action, there is an unavoidable stepping back process, which entails that any inductive knowledge of what we will, or would, do, or decide to do, in the natural course of things is knowledge of the situation that confronts us: and in respect of this situation, once recognized, it must be true (logically must) either that we intend to alter it in some way, or that we intend to leave it as it is and to allow events to take their natural course. That is, it must be true

that we have some secondary intention. And, if we do not take this secondary intention into account, we cannot know what we will do.

I. M. I'm not sure that I see what the 'unavoidable stepping back' is like, or in relation to what predictions it is supposed to take place. It's worth recalling that the predictions we do in fact make, whether in strictly scientific terms or not, about our conduct usually don't have any very high certainty status. It's difficult in practice to make exact predictions about complex individuals — and it may be difficult to isolate aspects of human behaviour in such a way as to make absolutely certain predictions about what one will do. Detailed and certain predictions of one's own future actions don't all that often occur. Usually when one speaks of the future, one is speaking of a tendency or a probability rather than of a certainty. We may often think it likely or even very likely that we'll do so-and-so, without being completely certain. If we put the matter in this way, it seems less obvious that a prediction *must* be a decision — for surely I can think it likely (but not certain) that I shall act in a particular way, and still be really undecided about how I shall act.

It looks as if Hampshire's thesis, if it is to have any compelling force, must concern certain, or rather (putting it from the subjective side) one hundred per cent confident predictions about my future acts. For it would indeed be odd for me to say both that I *knew* how I would act and that I was *undecided* how I would act. Something that looks like a conceptual necessity does seem to live here. One can't (logically) be undecided about what will happen

when one *knows* what will happen. I am not sure, however, how much we can get out of this particular logical necessity. Of course it remains an open question whether I *can* know what will happen — but Hampshire's point concerns my own *picture* of the future, *i.e.* what I think I know about what will happen. And here I would repeat that in fact we're not usually so very certain about what we will do.

P. L. G. I agree that it seems odd to suppose that a person could both claim on purely inductive grounds that he was *absolutely certain* what his decision would be when he eventually took it and at the same time profess to be genuinely undecided as to what to do. Perhaps a man may — following Pears's example — assert with complete assurance, and on the basis of past experience, that when the moment comes his decision will be determined by a different view of the situation, a different appraisal of what is involved, from that which he now has or makes. But to what extent can such a man be said to be *undecided*? For isn't it central to the concept of indecision that a man who is truly undecided about what to do about a practical problem must view it, not merely as one which he has to solve, but as one which he does not yet know *how* he will solve? It may be objected that, in the (admittedly rather unusual) type of case under consideration, the man is still 'undecided', at least in the sense that he has not yet endorsed the reasons which will, he is sure, finally sway him. But how convincing is this? Wouldn't it be more natural to regard him as being simply unhappy or apprehensive at the thought of his taking the decision in question — even as suffering

from a kind of forward-looking 'regret'? Thus, in *The Golden Bowl*, Henry James considers a man who, contemplating an action he firmly believes he will do, wishes he wasn't so sure he would do it: 'he knew himself, however, well enough not to doubt; he knew, coldly, quite bleakly, where he would, at the crisis, draw the line'. Would it be very plausible to characterize such a man *tout court* as undecided? But, if not, must we fall back on saying that in *some* sense he must really have already made up his mind? Doubtless there are many subtleties of interpretation conceivable here. For instance, it may be hard to treat some things we say about our future behaviour as being straightforwardly *either* 'pure' inductive self-predictions *or* 'pure' decisions. Again, I should not want to deny that many predictions about 'primary' decisions may involve implicit 'secondary' decisions: in this respect Hampshire's arguments are intellectually, as well as morally, compelling. I wonder, however, whether he does not want to apply his framework too rigidly. May there not be circumstances in which a man, with some justification, sees himself as a kind of 'prisoner of his personality', recognizing certain things about himself to be inescapable, certain of his choices to be inevitable? Of course it may be said that even here one can speak of 'decisions' — decisions to 'acquiesce', to 'resign oneself', and so on. But how far, I wonder, can one travel along this line without at some stage straining thought and language, and without harbouring the uncomfortable suspicion that one may be attempting to impose categories that do not really apply?

S. N. H. You are, I think, arguing that there are subtle mixed cases of knowledge of our own future actions, which cannot be too simply classified as clear cases of decision or as clear cases of inductive prediction. I recognized this in my book. The man in *The Golden Bowl*, who knew himself well enough not to doubt how he would act when the crisis came, is indeed such a mixed case: but still the knowledge of himself, which was the basis of his confidence, was not really inductive knowledge. It was nearer to a candid survey of his present dispositions and intentions. He *had* regretfully decided ultimately to draw the line: but at the same time he was, very self-consciously, stepping back to see the intention forming. He watched the process of intention forming in himself.

I. M. Suppose we get back to this matter of 'stepping back' and envisaging the situation. I think that Hampshire holds that beings who do this have freedom of the will. What *is* this 'stepping back'? If a person who is considering his own future course of action predicts what he will do, then either his prediction contains a decision to accept what he foresees, or else he can step back and make a secondary decision either to acquiesce in the predicted course or to resist it. And Hampshire would claim he has *got* to do one of these things. We can't be simply aware of ourselves as a sort of machine of tendencies, because the awareness constitutes an extra bit. This argument certainly seems attractive. It appeals to our own consciousness of ourselves as free, and it has something of the compulsive quality of the *cogito* of Descartes. But what does it mean

exactly? '*If* you clearly envisage your future situation and how you will reason and decide, you *must* be thought of as taking up a standpoint about it now.' Now even this isn't necessarily so, because one may be neutral about one's future reasoning. But what about persons who don't 'clearly envisage' — either because of muddle or because of some positive (perhaps religious) belief? Is the 'stepping back' unavoidable in their case? If Hampshire says it *is* unavoidable he's implicitly holding that a human being is always rational enough to know what he's doing; and this may be denied. If he says it is *not* unavoidable he's severely narrowing what we usually think of as the area of freedom.

D. F. P. Let us look at the mechanism of this argument. In the cases to which it is applied, all the agent's predictions go through his own reasoning and decision: *i.e.* to use the term that I introduced earlier, in all the predicted fragments of his biography, his reasoning and decisions are operative. Now we might ask whether the agent himself has to make the prediction about himself, or whether he can accept it when it is made by someone else. I think that the answer to this question is that Hampshire doesn't mind who originates the prediction, so long as the agent believes it: either way, it is a first person prediction.

So far, I think that everything is clear. But the next question of interpretation is more difficult to answer. What accounts for the alternative that is left open at each stage of the argument: for example, why does Hampshire say, at the first stage, that either the prediction contains a secondary decision by the

agent to acquiesce in the predicted course of his own biography, or else he can step back and, in the light of his prediction, now make a secondary decision, either to acquiesce or to resist? I think that the answer to this question is that the alternative depends on whether the agent regards the prediction as absolutely certain or not. If he regards it as absolutely certain then there is nothing more for him to do, and according to Hampshire his prediction in some sense contains a secondary decision: though in what sense it contains a secondary decision is, as Gardiner has pointed out, a very difficult question to settle. If, on the other hand, he does not regard it as absolutely certain, he can step back and make a secondary decision, one way or the other. And here it is of the utmost importance that this alternative does not turn on the question whether the prediction *is absolutely certain*, but only on the question whether *the agent regards it as absolutely certain.* For the agent might think it was not absolutely certain, when in fact it was absolutely certain.

P. L. G. It is perhaps worth stressing that the 'regress argument' only applies where the inductive prediction is either made or accepted by the agent himself. One can, for instance, suppose the prediction about what a person will do to have been made by someone else, who then keeps it to himself, subsequently observing that the agent acts in conformity with it. In such a case it is arguable that the issue concerning the agent's objective freedom remains where it was; it is not touched by the kind of reasoning we are considering.

What, though, of cases where the argument

applies? Certainly the sceptic about objective freedom might allow that a man may (seriously and in good faith) make a secondary *resolution* to try to alter the predicted course of his own reasoning and decision. Yet he might nevertheless claim that to draw the conclusion such a man can therefore always make a secondary *decision*, in the light of his prediction, whether to 'acquiesce or resist', may be in an important sense misleading. For the notion of decision, as commonly employed, often carries the implication that what is so classified will be (to some degree at least) effective or operative. Yet here — he may urge — there can be no *a priori* guarantee of efficacy. Whether in a given case a person's resolution to resist will be operative is an empirical question; one, moreover, on which the issue being discussed largely turns.

D. F. P. I agree. But there is also another, more complicated, reason why the regress argument does not establish objective freedom.

There are two quite distinct arguments that the determinist would use against objective freedom. He might argue, as you are suggesting, that the predicted action will be performed whatever the agent decides: *i.e.* the agent's decision is, according to him, inoperative. In that case, I agree that even when the agent is told the prediction, it will often make no difference to the outcome. He will perform the action anyway, whatever secondary decision he makes.

But sometimes the determinist will use his other argument. He will argue that though the decision is operative, it is itself determined. Now what

happens if the agent is told both that he will make a certain decision, and that because of it he will perform a certain action? This is a more complicated case.

He might, I suppose, try to make the opposite decision, if only because he was counter-suggestible. And then he might find that he simply could not make the opposite decision, or at least could not achieve anything that would count as a decision in this direction. Of course, he would not necessarily mind this: for he might very much want to perform the predicted action. Determinism does not imply reluctance and regret.

Alternatively, he might succeed in making the opposite decision, and in not performing the action. But this would not establish objective freedom. For there might be another prediction, unknown to him, which foretold that he would do just this as a result of being told the first prediction.

P. L. G. I am far from sure that Hampshire does in fact regard the sort of argument we have been considering as an argument for 'objective freedom': I think, however, that he does in some sense want to maintain that it establishes that it is impossible to escape the subjective burden of freedom. Here we could perhaps go back for a moment to the notion of the 'inoperative decision'. Suppose a man thinks it very likely he will do something, decides to interfere with the 'anticipated course of his own psychological history' and nevertheless ends by deciding and doing as he originally predicted he would? And suppose this happens, not once, but often in a particular recurrent type of situation?

May such experience not lead him to the conclusion that he really cannot, in this sphere of his conduct, 'help himself', his acceptance of this conclusion representing not so much some 'implicit' decision on his part, as the recognition of an unavoidable limitation upon his power of choice? It seems to me to be not only conceivable, but (in some circumstances) wholly natural for a person to look at and describe *particular* aspects of his personality and behaviour in this way; though it might, of course, be a different matter to suggest that a man could take such an attitude towards his character and conduct *in general.*

S. N. H. About the 'stepping back' and secondary decisions; any inductive knowledge that a man acquires of the determinants of his behaviour, which are other than his own decisions, changes the situation, as he confronts it; he sees that certain possibilities which he had believed to be open, are closed. But he now has to decide what he is to try to do in the new situation. He is not absolved from further decision, just because he now realizes that the primary decision would have been inoperative. This is the 'stepping back', which opens new possibilities of things that he might now try to do, given his knowledge of the natural course of things.

Consider an example of an inoperative decision. Suppose a hospital full of drug-addicts in conditions in which it is absolutely certain, as a well-established clinical fact, that they will all ask for the drug within twenty-four hours of the cure having begun. Suppose that several of them say that they have decided to ask for it, and then give varying reasons, carefully considered and plausible, for their decision: others

simply shout for it. Here it is rather strained to say that the first group made decisions, but inoperative ones, because they would have asked for the drug anyway. For, in such cases, it is not as though the so-called decision was itself perfectly all right, and the trouble lay entirely in the sequel, in the connection between the decision and the action. There is something wrong with the so-called decision itself: it is made against the wrong background, since the action is inevitable. So I think that if we are going to say that there is a decision in such cases, it would be better to call it a *void* decision. For it is very like the case in which a man is not empowered to make a public decision: it is *ultra vires*, and so nobody listens to him.

The determinist's claim is that we often go through the motions of deciding, and in good faith, when in fact our so-called decisions are void, because possibilities that we thought were open to us were not in fact open. This is the attack that I try to meet in the book by arguing that whenever a man predicts his future reasoning and the workings of his mind, he can always step back and make a secondary decision to interfere with the predicted course of these events. This, I maintained, is a conceptual necessity.

I. M. If we allow the existence of 'void decisions', we seem to have given the determinist a possible move which I don't think Hampshire has blocked. A decision is void when a man goes through the motions of deciding when he's not really in a position to decide. (The case of the drug-addict.) And I don't see that one could show other than negatively that any particular case was a case of a true and not

an apparent decision. The determinist will say, 'You agree that in some cases where you thought there were decisions, there really were not. Well, all cases are like that.' To which one can perhaps only reply that there *are* situations where we come to regard reasoning as a non-effective surface movement, but it hasn't been shown that *all* situations are like this.

Hampshire allows void decisions in cases like the drug-addict case, and might allow in less extreme cases that people 'drift', fail to think out the 'natural course of events', and so on. So the sense of freedom he's trying to prove is not necessarily exemplified in all human action. It looks more like an *achievement* than a necessary adjunct of being human. And Hampshire's 'conceptual necessity' would concern how we *have* to picture ourselves *if* we picture ourselves. This leaves open the possibility that we do not picture ourselves, *and* the possibility that the picture may be wrong. The logical argument will be simply: we can't picture ourselves as unfree. Yet, as we've already said, the drug-addict can. And we've envisaged less extreme examples where one may think of oneself explicitly as a prisoner of one's personality. Perhaps all one can say, and here I agree with Gardiner, is that it would be difficult to imagine extending the idea of oneself as a prisoner to cover the whole area of one's conduct.

D. F. P. But aren't you forgetting that a determinist will not maintain that all decisions are void and make no difference to the outcome? Certainly he will say that this is so in some cases. But in most cases he will use an entirely different argument; he

will argue that, though the decision is often not void and even essential to the outcome, it is itself determined. Now the regress argument simply has nothing to do with this part of the determinist's thesis.

I rather think that when Hampshire uses the word 'determinist' he is only thinking of the first part of the determinist's thesis — the part which says that some decisions are void: and when he says that a decision must be between possibilities that are really open, he is not raising the question whether or not the agent's psychological history develops freely; he is only insisting that, if it contains an event which can properly be called a decision, that event must be the source of whatever action he ultimately performs. But this leaves the more important half of the determinist's thesis untouched.

I. M. Hampshire seems to picture the mind like this: *reason*, giving us knowledge of tendencies, and *will*, a capacity to decide about tendencies, which gives us an alternative kind of knowledge of the future. He equates *self-knowledge* with the opening of possibilities before us in a manner independent of will. He refuses to allow that feelings might put pressure on reason in a way that couldn't be controlled by the will. For he wishes to keep reason intact and uninfluenced by anything outside itself. Or rather, if it is influenced by anything outside itself, perhaps by a particular emotion, he wants to say that reason can step back and make a new assessment of this total situation, and that in the light of this assessment the will can make a new decision. But is this always realistic? *Aren't* some decisions really predictions, predictions that we can't reason in certain ways, or

see things in certain ways, because our feelings make this impossible for us?

P. L. G. Yes, I agree that there are great complexities in the notions of knowing or understanding oneself, and also that there may be connected difficulties in the idea of there always being a detached or 'pure' power of rational choice, clearly separable from the deep-lying emotional or motivational tendencies which inform a person's nature. It would appear, in fact (at least if the impression given by many writers whose work is of a characteristically self-exploratory or introspective description is any guide) that certain kinds of self-knowledge may lead to a man's realizing some possibilities — not only of action or feeling, but also of 'reason' — to be closed to him. For example, he may recognize that forms of rational consideration or practical argument exist such that, whatever attraction they may on occasions seem to have, they could never in the end prove decisive with him, and that it would be simply self-deception for him to suppose otherwise.

One further, quite different and very general, point. Have we throughout centred perhaps too exclusively upon a particular interpretation of determinism? What, for instance, of the sort of considerations upon which the thoroughgoing determinist is possibly most likely to fasten — such as the influence of cerebral or glandular activity on human choice and behaviour? I am not clear how far Hampshire regards his thesis to be relevant to determinism so conceived, or if he considers it to be relevant to it at all.

S. N. H. First: objective determinism. I only intended to deny one of its alleged implications:

namely, that it leaves no place for an agent's decision between possibilities.

The reason why I deny it is this. As has been pointed out, the determinist offers us two possibilities: either he says decisions are void — as in the case of the drug addicts, since what they decide makes no difference to the outcome — or the decisions themselves can be accurately predicted with absolute certainty, so that, though they may not be void, they are determined in advance.

Now the truth in the determinist's thesis is that at any given time I may not know whether a decision of mine is determined or void. But the false part of his thesis is that he forgets that, whenever we think we know this, we can always step back and *try* to alter the initial conditions. Even in your suggested case, knowledge of the working of the nervous system would always put us in a position to *try* to alter its working. This is how my regress argument refutes the corollary that is often drawn from the determinist's thesis, viz. that there is no place for decision.

Iris Murdoch, in addition to sharing Gardiner's doubts about freewill and the regress argument, questions the basic distinction that runs throughout my book: between the self as trying and deciding on the one hand, and on the other, the natural course of things, which is to be known and understood. Is this hard dichotomy a conceptual necessity which everyone must accept? Or does it depend on one particular ideal of 'reason', a normative word — one, among others, which is being imposed by a persuasive definition? This is a very

difficult question, which touches on the nature of any philosophical scheme and therefore of philosophy itself. I am inclined to agree with Iris Murdoch that this sharp opposition between man's agency, and free decision, and his inductive self-knowledge and their alleged complementarity, is only one scheme among others; that this is a particular liberal ideal of individual freedom and activity, and that freedom and self-knowledge may, without logical contradiction, be differently defined, with the line between different kinds of knowing what one is going to do falling in different places. In any final analysis, no unique set of definitions of mental concepts is forced upon one as unavoidable. And this indeed I did say in *Thought and Action*.

7

POSTSCRIPT

DISCUSSIONS of freewill and of determinism naturally go together; these ideas seem like two sides of the same inscrutable coin. Yet when we look more closely, it emerges both that the celebrated opposition is less clear and definite than at first appears, and that there is much in the area of freewill that demands discussion and clarification in its own right. Both of these points are illustrated, in particular, by Chapters 2 and 3 of the present book. Mr. Pears, Mr. Thomson and Mrs. Warnock in their discussion of the will itself, and, again, Professor Hart in his analysis of legal responsibility, are indeed starting from notions that have frequently been invoked in the opposition of freedom to determinism, namely the notions of efforts and acts of will. Their discussions show, among other things, that these notions are complex and problematical, and at least less generally applicable than freewill theorists have hoped. Professor Hart further shows that if one tries to use, in the practical contexts of the law, the idea of an act of will or volition as a criterion of responsible action, nothing really comes of it: the supposed criterion emerges as a misleadingly positive stand-in for the variegated criteria that are actually applied in deciding that certain actions are *not* responsible.

So there is much to be said about freedom and responsibility that is not directly related to determinism. Nevertheless, the relation to determinism remains the heart of the problem of freewill, and it is to determinism and its surroundings that most of these concluding remarks will be directed. Chapters 4, 5 and 6 of this book have raised a very complex set of issues on this topic, and a number of challenging arguments; and it seemed appropriate that some concluding attempt should be made to set some of these out more formally than was possible in the limit of broadcast discussions, to relate them one to another, and to draw out some further implications. This I shall try to do in this Postscript. In doing so, however, I fear that I inevitably exceed my brief to some extent, and produce criticisms and arguments of my own, which go beyond the texts I am seeking to weave together. Here I must offer a general and preliminary apology to the contributors to the earlier chapters, both for the use and misuse I make of their views, and for various criticisms to which, by reason of my safe concluding position, they have no opportunity to reply within this book.

In the discussion that forms Chapter 4 of this book, there seem to me to be three main lines of argument that come up against the idea that a deterministic science of human behaviour could undermine our ordinary concepts of human action and responsibility. These are, though not in the order in which they occur in the discussion, first, that determinism in general terms is not so much a thesis as a programme: second, that allowing that some form of determinism could be coherently

formulated, then if it were *physical* determinism — *i.e.* a theory couched in purely physical or mechanical terms — it would not contain or imply any descriptions of human actions at all, and so could not impinge on them; and third, if this last difficulty were overcome by having some *psychological* form of determinism, which did refer to human actions, it nevertheless could not displace or upset our normal explanations of such actions in terms of the agent's motives, reasons, intentions, etc. — and this (I think it is implied) is enough to show that our ordinary concepts of responsibility would still be applicable.

I should like to say something about these three lines of argument in turn. First, the assertion that determinism is a programme rather than a thesis is itself ambiguous. If determinism merely claims that for every state of affairs some explanation can be found in some terms or other, then indeed it is only a programme, in the sense that no determinate thesis has been stated at all. However, determinism can be taken to claim that every state of affairs can be explained by reference to other states of affairs together with a set of natural laws, all of this to be expressed in terms of *some specific type* of descriptions; and if the determinism is to be universal, it seems that these would have to be purely physical descriptions. Determinism in this form, while in one sense, it still states only a programme, viz. a programme of physical researches to make good the claim, *also* states a thesis. It might be objected that such a thesis was vacuous, because not falsifiable: however often one failed to find physicalist explanations

satisfying the requirements of the thesis, one could still claim that they were there to be discovered. However, this objection could at least not be well combined with the second strain of argument I am considering, which states that human actions, by not being merely patterns of movements, would be isolated from such a physicalist determinism. For the second objection admits that most actions (at least) *involve* physical movements or changes; and there are further physical states of affairs that are evidently the results of actions. If the second objection is correct, none of those movements or states of affairs can be fully explained in the way the physical determinist requires; unless actions and bodily movements are explanatorily quite independent of each other, which is surely incredible (see Mr. Thompson's point, p. 65).

There is another, and related way, in which this determinist thesis is substantial. Being a purely physicalist theory, it would have to exclude teleological patterns of explanation, that is to say, explanations in terms of purpose: since it is a feature of developed physical theories that they have no room for teleological explanations. Thus determinism of this type involves the claim that the teleological explanations with which we are familiar in the case of human and indeed other animal behaviour, should be replaceable by explanations in terms of initial conditions, stimuli and natural laws. The claim that this is possible is certainly substantial, and there has been a good deal of controversy in recent psychological literature on the question of how, if at all, the programme is to be carried out.

These consequences of the determinist view would become explicit if it were stated in the form of saying that every state of affairs can be explained in terms of an *earlier* state of affairs.

There is, admittedly, an element of vagueness in this formulation of determinism, which lies in the phrase 'can be explained': this does not reveal exactly how strong a requirement is laid on the determinist in carrying out his programme. It would indeed be clearer what the requirements were, if the determinist claimed that every state of affairs could in principle be *predicted* from earlier states of affairs together with the natural laws, and the classical determinism of Laplace did take just this form. However, while this formulation is clearer, it also seems to be too strong; it seems to involve implicitly the notion of a description of the total state of the universe, from which alone, it might seem, categorical and certain predictions could be derived, and it is far from clear that any sense can be given to this notion. Most scientists would undoubtedly be content with something less than this vast Laplacean claim; while predictive power remains an essential part of scientific explanations, the predictions can be offered in most cases only within certain limits of accuracy, and subject to certain sorts of reservations. To decide exactly how much can sensibly be required on such issues is a complex question of the philosophy of science. For the problem of freewill, however, it might be suggested that it would be enough to mark someone as a physical determinist if he claimed that human behaviour was in principle explicable in exactly the same terms, according to the same

principles, and to the same limits of accuracy, as the behaviour of purely physical structures of a comparable degree of complexity. Such a claim would seem to be contentful, and to raise the problem of freewill; even though it eschews the ambitions of total Laplacean determinism, and even though, of course, the physical theories it promises remain at this moment a programme rather than a realized fact.

To turn now to the second line of argument about physical determinism: this was that it cannot reach up to, or embrace, human action, because human actions are more than, or different from, mere agglomerations of physical movements. Mr. Warnock has shown some of the difficulties about this argument in Chapter 5: his point, that even if descriptions of human actions as actions are not equivalent to descriptions of them as physical events, certain descriptions of the first type are at least incompatible with certain of the second type, is an evidently forceful objection to any very general use of this argument. I wonder, however, whether his line of objection could not be carried rather further. The argument in question runs: human actions consist in some part of physical movements, but no description of the physical movements adds up to a description of the action; physical determinism is concerned only with physical movements; therefore it can ever produce any statements which count as descriptions of human actions. The basic mistake in this argument is surely the supposition that it is *these* physical movements — the gross bodily movements — that the physical determinist would alone con-

sider.[1] The route of the physical determinist's explanations would surely have to run via neurophysiology: it would be the states of the brain and the nervous system with which he would be concerned. Here he would have a much wider range of information to explain and predict human behaviour in its relation to its environment than mere gross bodily movements. For instance, there would be elements in the theory that corresponded to what we mean when we speak of such movements as being purposive, even though those elements would not themselves be of a purposive structure, but in terms of initial conditions and stimuli.

Of course, it is true that the determinist could not find, at the level of gross bodily movements, anything to explain these factors that make actions more than mere bodily movements. But it is not at that level that he hopes to find such explanations. This also answers the specific point made by Mr. Strawson (p. 66), that the same physical movement may be a part of quite different actions, and conversely. The physical determinist is not looking for a one-to-one correlation of actions and bodily movements, and the idea that he needs to is clearly gratuitous.

However, it may be objected that a neurophysiological account would still not entail any statement which was itself a description of a human action: so the determinist programme would still not have been carried out for human actions. But against

[1] It is surely only by restricting his attention to this kind that Mr. Strawson could have come to make the claim (p. 66) that 'there is no question of such correlations ever being established'.

this, one may consider the case of a computer. It seems clear that a computer is a deterministic physical mechanism, yet we do in fact describe some of what it does — as we should now say, by analogy — in terms of such activities as calculating, deciding, checking, learning — even perhaps, hesitating or plotting. Now no statement that uses such terms is contained in, or entailed by, the physical theory that explains the design and workings of the computer. Yet it is evidently true that all these ascriptions of 'activities' to the computer actually apply to states of it which can be expressed in the physical theory, and that correspondingly all the 'activities' can be satisfactorily explained in terms of the physical theory; and these points, which are the significant ones, still hold even if we allow that we apply some, at least, of these descriptions to computers because *we* have designed them to do certain jobs, and have given them purposes, as it were, from outside. It would seem that the determinist would have done enough, if he could find computer-type physical descriptions and explanations similarly related to human actions. The point that descriptions of human actions, as such, could not form part of any physical theory, is not enough *by itself* to show that he could not do this.

Here it will be said [1] that computers, unlike human beings, do not have consciousness. The idea of 'consciousness' is the seat of notorious and difficult problems. But it is at least clear that to appeal to the existence of human consciousness would not be

[1] Though it is not in Chapter 4. I am pursuing a familiar line not considered in the discussion.

to reinforce the argument under discussion, but to produce a new one. The argument under discussion appealed merely to the fact that descriptions of physical movements systematically fall short of descriptions of human actions as such. An appeal to the existence of consciousness in this connection, however, would have to claim that they do not merely fall short, but fall short in a particular way, namely by leaving out certain factors of consciousness: it would claim, for instance, not merely that writing one's will was more than moving one's arm in such-and-such ways, but that it was more than this because it presupposed further the occurrence of certain sorts of thought, etc. This would be a different argument.

The heart of this new argument would, I think, lie in two considerations. The first would be that conscious events or states seem to be *different* events and states from any physical (*e.g.* neurophysiological) correlates that they might have; because it seems clear that we can be aware of such events and states, and certain of their occurrence, without knowing anything at all about such physical happenings (we do not inspect our brains to find out that we are in pain, or thinking about yesterday's interview). Second, the descriptions of these conscious events and states could not form part of any purely physical theory. Hence it would seem that there was a class of distinct events and states, of central significance to human life, which could not itself be incorporated into the physicalist theory; and while it might be claimed that all such events and states were contingently connected with, or correlated with, physical events and states that could be incorporated into

the theory, one would be left with the apparent anomaly that the contingent connection or correlation could not itself be stated in terms of the purely physical theory, since in order to state a correlation in terms of a given theory, I must be able to describe *both* the correlates in terms of that theory.

Here we do seem to have a genuine difficulty: that systematic correlations between conscious or psychological states on the one hand, and physical states on the other (the theory of 'psycho-physical parallelism', as it is sometimes called), seems the best that the physical determinist could hope for; yet to accept it as ultimate would seem to impair the purity of his enterprise. There are some very old problems here still unresolved.

One way of getting round such problems for the determinist might be for him to abandon the requirement that determinism should be physicalist: he might admit a psychological determinism, not ultimately reducible in terms of a physical theory. This would certainly weaken one aspiration of some determinists, to produce a *universal* system of explanation; but perhaps there would be enough left for determinism without that. The subject of psychological determinism comes up in the third of the arguments I have quoted from Chapter 4. The form of psychological determinism being considered is not one suggested by theories of behavioural or experimental psychology — which indeed, in many cases at least, seem to aspire to a physicalist basis, and hence not to be very ultimately psychological — but rather psycho-analytical determinism, based on considerations about the unconscious mind. The

issues that arise in this connection are rather different from those discussed elsewhere in this book, and in the earlier parts of this Postscript. Psycho-analytical explanations are, for one thing, very low in *predictive* content, and in some ways more like explanations in history than those of a natural science. Again, there is no question, at least with psycho-analysis as it stands, of replacing teleological explanations of human behaviour with non-teleological ones. On the contrary, psycho-analytical explanations are themselves characteristically purposive; not only are certain (unconscious) motives and purposes ascribed to the patient, but also certain mechanisms of the mind are themselves described purposively, as, for instance, when it is said in the Freudian theory of dreams that the unconscious adopts certain symbolic devices to evade the activities of the censor. It may be that these latter descriptions are intended only figuratively, and in principle could be replaced; but the former teleological elements, at least, seem essential to psycho-analytical theories. Such theories do not replace teleological explanations by non-teleological ones, but replace or supplement our ordinary, superficial, teleological explanations by deeper ones.

These features of psycho-analysis often appear to the natural scientist, and would certainly appear to the physical determinist, as weaknesses of it. But these weaknesses, if such they are, have compensations, in that the theories do actually apply to complex forms of human conduct, and explanations of human conduct are actually produced, instead of being (as the physicalist's still are) hopes for the

future of science. Their influence on actual practical thought about freedom and responsibility is correspondingly greater. This is to be seen at both a particular and a general level: in the particular, when psycho-analytical explanations of certain sorts of aberrant behaviour carry conviction and modify the tendency to ascribe responsibility and blame; in the general, with the thought that perhaps *all* behaviour is unconsciously determined in such a way as to defeat responsibility.

In Chapter 4, the participants in the discussion approach these issues from the direction of asking what effect psycho-analytical explanations have on the status of the ordinary conscious reasons that people give for their actions. A distinction is drawn between, on the one hand, explaining why a certain reason *is* a man's reason for acting in a certain way, and, on the other hand, showing that the reason was not a reason, or at least not a sufficient reason, at all. It is suggested, I think, that psycho-analysis might in general do the first, but that this would not undermine our ordinary concepts of human action and responsibility; the second might undermine those concepts, but is something that psycho-analysis cannot *in general* do. The only case in which psycho-analytical explanations would seem to undermine our ordinary reasons for actions is where the reasons are bad reasons, in particular, rationalizations; and here it is not so much that the explanation does the undermining, as that the reason has no real foundations anyway (we do not believe that this *was* the man's reason for action), and the explanation shows what in fact it was that governed his actions. But

this particular sort of situation could not conceivably be generalized, to give rise to the general doubts about responsibility; for it is absurd to suggest that all our ordinary reasons for acting are bad or inadequate, and it is only when they are that the explanations are in place.

There are a number of complex issues involved in this argument. First of all, the participants in the discussion appear to be agreed, at least after Mr. Thomson's spirited remarks (p. 56), that the argument cannot be based merely on a general thesis about explanation, to the effect that we look for explanations only in unusual cases, that the normal *needs* no explanation. This thesis has, surprisingly, had a certain currency in recent philosophy, but is evidently absurd: it rests on an idea of explanation which may serve the purposes of a jobbing electrician, but is happily not shared by natural scientists, or by children. So the argument must rest on some special features of the explanation *of human actions*; and one special feature seems to be, that human beings give reasons for their actions, and those reasons are in normal cases adequate. So any thesis which implies that the reasons given in normal cases are inadequate must be a false thesis. But any 'undermining' generalization of psycho-analytical types of explanation from the cases of rationalization to all cases would involve just this consequence, and hence must be mistaken.

The difficulty here is with the expression 'adequate' reasons. One important point about this expression is brought out by Mr. Strawson (p. 61), in the form of a distinction between '*his* reasons' and '*the* reasons', the basic consideration being whether

the issue is merely one of understanding how an agent came to do a certain thing, or of his showing that it was a reasonable, sensible, rational, etc., thing to do. I think this might be put as a distinction between reasons as explanatory and reasons as justificatory. It is not altogether clear, however, how Mr. Strawson would want to apply this distinction to the present issue: which sense of 'adequate reason' is the one in which it is absurd to deny that normal reasons are adequate? Which of them is particularly connected with the ideas of freedom and responsibility? At one point, Mr. Strawson says something that suggests that it is the second, or justificatory, sense of 'adequate' that is at issue here: 'his standpoint here [*i.e.* in giving a justificatory reason] is that of an agent who is, or wants to be thought, reasonable and responsible'. This *might* be taken to imply that it was only action that had adequate reasons in the justificatory sense that was responsible. If one pressed this further — further, I am sure, than Mr. Strawson intended — one might end up with consequences of a very unacceptable kind, to the effect that it is only actions that accord with norms of justification that are to be regarded as free. But leaving such distant, and perhaps avoidable, consequences on one side, there is in any case an objection to taking *this* sense of 'adequacy' as the one appropriate to refuting the generalization of psychoanalytical considerations. For that generalization consisted in generalizing the notion of a rationalization, and it is not in this, justificatory, sense of 'adequacy' that a rationalization is an inadequate reason. On the contrary: a rationalization is a

reason precisely invented to satisfy the requirements of justification, to make the agent's conduct look more reasonable or rational than it is. What makes it a rationalization is that it was not *really* the agent's reason for his conduct, *i.e.* that it is inadequate in the explanatory sense.

Is it obvious that in the explanatory sense of 'adequacy', it makes no sense to suggest that our normal reasons are inadequate? In one way, it does seem obvious: since the cases of rationalization are precisely distinguished from normal cases by the explanatory inadequacy of the reasons given, it would seem that explanatory adequacy in the other, and normal, cases was well enough established. But here, I think, some further distinction needs to be made. There is more than one thing that might be meant by saying that a reason was explanatorily inadequate. We might mean that the stated reason failed to correspond to the actual nature of the agent's action or policy, that the correct account of what he was up to would leave out the considerations mentioned in his reason altogether — except perhaps for stating that they occurred as a rationalization. This will be the case, for instance, with a man who refuses to employ a person because (as he states) he appears inefficient, but actually (we believe) because he is Jewish, the employer being anti-Semitic. Here the inadequacy of the reason lies in this, that the man's real policy is that of anti-Semitism, and inefficiency does not really come into it, except as a cover. In this sense of inadequacy, it is obviously untrue that all the reasons normally regarded as adequate are in fact inadequate.

However, we might mean by saying that the man's reasons were explanatorily adequate, that they, or the decision for which they were the reasons, give the complete story about how the man came to act in the way in which he did. In this sense, so far from its being the case that reasons normally given are adequate, they scarcely ever are: they all operate against a background of dispositions, desires, etc., which would have to be further mentioned in giving a complete story. It seems open to the psychoanalytical determinist to claim, first, that this background contains highly important elements of which the agent is not aware, and second, that these elements were laid down by early experiences which he has forgotten and which were not in his control. Now to claim this would be not to *undermine* the ordinary professed reasons, as the professed reasons are undermined in the first, or rationalization, sense of 'inadequacy', and this is, perhaps, just Mr. Strawson's point. Nevertheless, such considerations may make us look at the ordinary reasons in a new, and perhaps slightly disturbing, light. They might suggest that the essential difference between rationalizations and ordinary reasons was a difference, in a limited sphere, of *self-knowledge*: that the rationalizer, who is at the conscious level sincere, does not know what he is doing, whereas the man who gives a 'real' reason does, to that extent, know. And this picture seems different from the one ordinarily associated with this distinction, which is rather a picture connected with *the will*: that in ordinary cases, we freely decide, but in cases of unconscious rationalization we are controlled by dispositions which are out of our reach.

In the new picture, we are in both cases controlled in good part by dispositions out of our reach, but when our reasons are adequate, in the sense of not being rationalizations, we do, to a limited extent, know where we are going. There is, indeed, something undoubtedly wrong with both these pictures. But to dispose of them would certainly involve going a long way beyond the consideration that there is some sense in which our ordinary reasons cannot all be inadequate.

However, it may be that it is in any case a mistake to think of psycho-analytical theory as underwriting some general and inescapable form of determinism. Some accounts, at least, of the therapeutic processes of psycho-analysis represent them as 'setting a man free' from unconscious motivations, and so allowing him genuinely and substantially to decide his future. If this representation is correct, psycho-analytical theory is self-limiting, as it were, and would give no ground for any general doctrine of determinism and absence of freedom. It would rather be a theory about *degrees* of freedom, and suggest factors which, unknown to us, limit our freedom, the removal of which would increase our area of free choice.

It is such a theory, I think, that Professor Hampshire is in some part concerned with in his work *Thought and Action*, which is the subject of the discussion in Chapter 6 of this book. Before turning to this, however, something must be said briefly about the relations of unconscious motivation to responsibility. A problem about this is posed, very briefly, by Mr. Warnock in Chapter 4, where he says (p. 62): 'When a man is brought to acknowledge an unacknowledged reason as his, he certainly sees his

action in a different light; but not as an action for which he can disclaim responsibility.' This remark, if true, would have a very radical effect on the issues under discussion. Since it claims, at least implicitly, that the fact that an action is unconsciously motivated does not affect the agent's responsibility for it, it must imply one of two things: either that unconscious motivation does not make an act unfree, or that an act's being unfree does not affect a man's responsibility for it. Of these, the second detaches responsibility for freedom, and thus denies the presupposition of much discussion in this book; while the first rules out any connection between unconscious motivations and lack of freedom, and hence gets rid of any problem about psycho-analytical determinism straight away.

But is Mr. Warnock's statement true? There does, indeed, seem to be something in it, but I suspect that what is in it is peculiarly connected with his considering the case of a man's regarding *himself* as responsible. It is an important fact that men are often prepared to regard themselves as responsible for actions, or results of actions, for which others do not regard them as responsible. Thus a man who has injured someone in a road accident through no fault of his own, may be dogged by feelings of remorse, attempt to make reparation, etc.; rather differently, an alcoholic who may be clinically beyond the point of unaided return may yet reproach himself for his continued drinking. Nor are such reactions necessarily regarded as neurotic, irrational, or even (in the first sort of case, particularly) unadmirable. Now it may be a necessary condition of

such reactions not being irrational that the agent himself should not be completely convinced that he was not responsible, that he should think that there was something he could have done to prevent the outcome. But the difference between the agent's own viewpoint and that of others remains, in that others may be quite convinced that there was nothing he could have done, or that there was nothing he could reasonably be expected to have done, and hence hold him not responsible.

So, I think, with unconscious motivations. Others, convinced that the kleptomaniac steals from unconscious motives, may hold him or her not responsible; but the kleptomaniac may, even when confronted with the facts of unconscious motivation, still regard himself as responsible, both in the more superficial sense that he thinks some reparation appropriate for the stealing, and in the deeper sense that he may feel that there was something he could have done to divert himself from the stealing. One root of this may just be a basic human drive to hold on to responsibility for as long as possible. But, of course, the agent may come to recognize that there is nothing he can do, unaided, to resist his tendency to steal; when this happens, he himself has granted a connection between unconscious motivation and lack of responsibility, which has already been applied by others to his case. Thus, while there are difficulties here (for different considerations may apply to different sorts of unconscious motivation), I do not think Mr. Warnock's claim, or implied claim, that unconscious motivation does not affect responsibility is really true.

The question of the agent's and of the spectator's viewpoints arises also, in a related way, with Professor Hampshire's views. But it is first necessary to go back a bit. I have already said that Hampshire's theory is connected with *degrees* of freedom. In *Thought and Action* he holds, if I understand him, that the degree of a man's freedom of choice is determined by the range of alternatives that he truly believes to be open to him. This has two aspects. A man's freedom may be limited by the fact that there are relevant courses of action open to him which he has not thought of, or believes not to be open to him; thus, a man's freedom may be increased by an increase of his knowledge and of his reflective powers. Alternatively, a man's freedom may be limited, in a rather different way, by his falsely believing courses of action to be open to him which are not. If this is so, the man's actions will characteristically fail to match his declarations of intention — he will fail to carry out the projects which he consciously adopts. Such a situation Hampshire regards as the central case of a limitation on freedom, in the sense with which he is chiefly concerned. Here again, an increase of knowledge may increase the agent's freedom; if he has some further end, to which he was pursuing the impossible courses of action as means, the recognition of their impossibility puts him in a position to devise some more effective strategy; or, more generally, he may abandon courses of action now recognized to be hopeless and try something else to which he also attaches value. There is, indeed, one special case, of the psycho-analytical type, in which the recognition of the previous limitation

of his powers may itself remove the limitation. It might be that a man was unable to establish satisfactory relationships with women because of some unconscious motivation — that he was, for instance, always seeking someone like his mother. Here the recognition, through therapy, that this is what he was seeking, *might* remove the neurotic motivation altogether, and so the limitation on his ability to establish a satisfactory relationship.

This part of Hampshire's theory, concerned with the degrees of, and correspondingly the limitations on, freedom, is not stressed in Chapter 6. What is there under discussion is an *absolute* sense of 'freedom' which goes with, and is presupposed by, these views. This absolute sense is that in which (Hampshire claims) an agent confronted with any statement about what he can or cannot do, can always step back and review these data as, at most, part of the frontier of limitations within which he can decide on a course of action. An agent can always regard his powers and his inabilities as a framework, like parts of the natural order external to himself, within which he can decide what to do.

But, it may be asked, are we certain that any real freedom is contained in this idea? For might it not be the case, if, for instance, determinism were true, that an agent's inabilities exhausted all courses of action except one: that what he could do, and what he would do, were one and the same? If this were so, the reflective power which Hampshire is stressing would not secure anything: in reviewing what was open to him, the agent would be forced to the conclusion that there was nothing he could do,

except what he was going to do. Hampshire's view is that this situation cannot arise, or at least cannot in general arise. For if an agent is confronted with a particular prediction about some specific action of his, as opposed to a general statement about his capabilities (and this distinction is very important for Hampshire's view), he can always 'step back', in the same sense as before, and decide to try to do something else, *i.e.* falsify the prediction. Two points are being made here, an empirical one, and a logical one. The empirical one is that very often he can succeed in falsifying it, and this is scarcely open to doubt: there are many things we do which are such that if it were predicted that we were going to do them, and we tried to falsify the prediction, we should succeed. The logical point is that, quite apart from this last point, if the agent himself announces a particular 'prediction' about his future action, whether derived from someone else's researches or his own, this cannot have the status of a prediction, but is rather to be taken as a *decision*: a decision, in this case, to let things take their course, and not to try to interfere with them.

For Hampshire, both predictions and decisions can constitute knowledge about an agent's future, but the knowledge is quite differently arrived at. With a prediction, it is theoretical knowledge, based on theory and evidence; with a decision, it is 'practical' knowledge, based on reviewing reasons for acting. It is his claim that any singular statement made by an agent about his own future must be regarded as the latter: even if its content is the same as that of some prediction arrived at on empirical

grounds, the decision element must always (logically) be there, in this case in the negative form of acquiescence.

A number of subtle and complex criticisms of these views are woven into the discussion in Chapter 6 — I cannot hope to review them adequately. It may be useful, however, to state what seem to me to be the main lines of objection:

(1) The alleged logical point is disputed, and it is claimed that a man can predict, without deciding, his own future (Gardiner, p. 82, Pears, p. 84, Miss Murdoch, p. 86; and elsewhere).

(2) The account of decision in terms of *knowledge* is questioned (Miss Murdoch, p. 90, cf. Pears, p. 95).

(3) It is claimed that the most that Hampshire's argument could establish was subjective, and not objective, freedom (Gardiner, p. 95; Pears, p. 96; and elsewhere).

(4) Miss Murdoch in particular claims that the 'stepping back' to which Hampshire refers, is not unavoidable, and that Hampshire's emphasis on it seems to contain some evaluative element (pp. 93, 100); this she regards as part of a more general, also evaluative, distinction sharply drawn between reason and will (p. 101).

To comment on these briefly: on (2), to take that first, it seems to me that Hampshire must qualify his thesis. For knowledge standardly implies truth, and the fact that a man has decided to do *X* by no means implies that he will do *X*. It does not even imply that he will try to do *X*, since he may, for instance,

forget his decision. The most that Hampshire can say is, not that one who has decided to do *X* does know that he will do it, but that one who says that he has decided to do *X* thereby *claims to know* that he will do it. Even this is perhaps too strong. For it seems that if a man claims to know a certain thing about the future, and also knows that something else in the future is a necessary condition of that thing coming about, he must in consistency claim to know that the latter thing will also come about. But it is a necessary condition of my doing, or trying to do, anything tomorrow that I should be alive tomorrow; and must a man who announces his decision to do a certain thing tomorrow in consistency claim to *know* that he will be alive tomorrow? He must indeed claim to believe it, but he might claim no knowledge or certainty in the matter (it would, I think, be foolish of him to do so). It might seem from this that Hampshire's thesis would be better put in terms of belief rather than of knowledge; but there are difficulties against this.

One such difficulty would arise from Hampshire's answer to objection (1). This objection is couched particularly in terms of a man's thoughts about a future decision, and it is claimed that the fact that a man knows how he is going to decide in a particular matter — and knows, *ceteris paribus*, how he will act — does not mean, as would seem to follow from Hampshire's thesis, that he has already decided. Hampshire's answer to this appears to be that while accepting that there are mixed cases of the kind Gardiner describes, he claims that a man who really *knew* how he would decide had, indeed, already

decided; for what would be left to this later process of decision? This point might be illustrated, by contrast, by reference to formal decisions of committees, etc., or of a particular person acting in an official capacity. Here, indeed, the members of the committee, or the official, may know how they will decide, without (in the official sense) having yet decided; but this is because deciding in that latter sense is a special procedure which has to be carried out in a particular context, according to particular rules. Even in this case, we might say that the matter had *really* been decided already: all that could happen later was that the decision was made official. In personal life there is in general no analogue of 'making the decision official', or of deciding in the official sense. In the absence of such an analogue, it is obscure what process a man would be referring to who said that he was going to decide tomorrow, but that he already knew how he would decide. What would he be waiting until tomorrow to do?

If this is so, there seems to be something in Hampshire's point. Nevertheless, it would be hard to sustain it without qualification. There are certainly some cases in ordinary life in which a man would be saying something quite intelligible who claimed that he had not yet decided, but that he knew how he was going to decide. The deciding here referred to might involve, for instance, a final review of the relevant considerations before taking an irrevocable step on the course of action; a man might intelligibly say that he had not yet conducted such a review, was going to do so tomorrow, but yet that he knew what the result of it would be.

Here Hampshire might still say that there would be something odd or unintelligible about such a man's claiming in these circumstances that he was *absolutely certain* about the results of the review, and consequently about his future action; for what would be the purpose of the review? To call the review of the considerations by the name 'deciding', is precisely to imply that *it* will decide what the man is going to do -- and if he is already certain on that issue, what purpose is left to it? It would, I think, be only with reference to a case in which absolute certainty was claimed that Hampshire could sustain his point. Suppose his point were granted for such cases; what is going to be the structure of his theory as a whole? Hampshire's claim that it is unintelligible for a man to say both that he is certain how he will decide, and hence how he will act in a particular matter, and that he has not yet decided, was a contribution to his analysis of decision in terms of certainty: the reason why my being certain about my future actions is incompatible with my not having decided is that such certainty entails having decided (if only acquiescently). But Hampshire, who is offering an *analysis* of decision, requires more than this; he requires also that having decided entails being certain. Difficulties about this latter point have already been seen in connection with objection (2); and in that connection, it was suggested that an analysis in terms of belief was the most that Hampshire could hope for.

Now we see a difficulty, in turn, about this suggestion. The analysis in terms of belief would claim a logical connection in both directions between my

decision and my belief about my future action: in its simplest form, it would claim both that my having decided to do *X* entailed my believing that I would do *X*, and conversely. But if the second, converse, entailment held, it would follow that it was unintelligible, or implicitly self-contradictory, to say that I believed that I was going to decide to do *X*, but that I had not yet decided to do *X*: the first part of this would imply that I believed I was going to do *X*, and this, by the suggested analysis, would imply that I had decided to do *X*. But it is quite obvious that there is nothing unintelligible or self-contradictory at all in saying that I believe that I am going to decide to do *X*, but have not yet decided to do so — the foregoing arguments have shown that if there is any incompatibility of this form, it must be in terms of certainty.

More investigation is needed into the relations of the concepts involved here. In particular, it is important that there are *degrees* of certainty about future actions, both on the side of prediction and on the side of intention and decision; objections (1) and (2) both suggest that Hampshire's thesis, which concerns the relation between these two sides, requires working out in terms of such degrees.

Objection (3) raises issues more radical than any of the others. According to this objection, Hampshire's considerations about the agent's 'stepping back' from any prediction about a particular future action of his, and deciding either to acquiesce in it or try to do something which would falsify it, cannot suffice to establish 'objective freedom', that is to say, to refute determinism. For either he could not

falsify it, in which case the prediction stands; or at least, there could always be another hypothetical prediction in the hands of an observer which stated (correctly) that if the agent were apprised of the first prediction, he would make a decision either acquiescing in it, or contrary to it, whichever the case may be.

Now the determinist who was seeking to defend his thesis against Hampshire would clearly have to rest his case in general on the second of these two possibilities, and not on the first. To hold that the first would always obtain, *i.e.* that the agent could never falsify the prediction whatever he tried to do, would not be determinism, but *fatalism*, and is obviously false, denying the evident empirical fact mentioned earlier. The second line, however, seems better; and indeed it is difficult to see how any determinist could possibly deny that among the stimuli of actions were remarks made to the agent, in particular supposed predictions about the agent's future action, the hearing of which could affect what he was going to do.

The difficulty is that the higher-order prediction thus invoked by the determinist seems to be hypothetical, merely stating what the agent would do if apprised of the first, lower-order, prediction. But it seems that what the determinist really wants are predictions that are categorical, and also, of course, correct. All he has in this situation is one prediction (the original one) which is categorical but probably not correct, and another (the higher-order one) which may be correct, but is not categorical. What he needs to remedy this state of affairs are two further predictions, one stating that the agent will be

apprised of the first prediction, and another stating that he will not be apprised of the second, higher-order, prediction: this latter being necessary to prevent the whole process starting off again. Where are these to come from? Here we seem to get back, by some route or other, eventually to the determinist himself, assuming him to be the scientist who is producing these various predictions; and to him, eventually, Hampshire's considerations will once more apply. For his certainty that he will or will not apprise the agent of these various predictions must once more have the status of a decision, and not a prediction. Nor would it avoid this to shuffle off all the predictions into the inside of some vast computer, churning out predictions independently of human knowledge or interference; for here again there would be the decision implied of not inspecting the computer and announcing its results. (Here one might say, indeed, that there was not only a decision to be made, but a temptation to be resisted.)

This line of argument, which is by no means new in discussions of determinism, seems to fit in with Hampshire's approach, and to show that his considerations will at least always break out *somewhere*. Whether this is enough for 'objective' freedom, I am doubtful. But it does seem to suggest something in this respect, over and above the inescapability of 'the subjective burden of freedom', as Gardiner puts it (p. 97). It suggests that a deterministic system of human behaviour would in principle be uncompletable, and if this conclusion is correct, it is certainly important. It suggests also, more broadly, a significant line of thought about the scientific process:

that this process and its results are themselves parts of human life, and that scientific predictions cannot be regarded as standing impersonally detached from human activities and commenting on them — for there is necessarily a place for human decision about what is to be *done* with such predictions. There seems to be enough in this line of thought to modify seriously any picture we might be tempted to have of a deterministic universe in which man was merely embedded.

Lastly, what of the 'subjective burden of freedom'? Miss Murdoch's criticism (4) suggests that Hampshire exaggerates in saying that the 'stepping back' to which he refers (which perhaps contains this subjective consciousness of freedom) is unavoidable; and she suspects an evaluative element in his views here. I think there are two different difficulties in this area, and they both centre on the question of how much psychological content Hampshire wishes to put into his notion of 'stepping back'. It will be recalled that on Hampshire's view, any statement made by an agent about what he will do in a particular circumstance has *logically* the status of a decision, even if the content be the same as that of some prediction which he or someone else has arrived at from evidence. Now if the phrase 'stepping back' is used in such a way that a man who announces a decision to do X, has *necessarily* 'stepped back' from any prediction previously presented to him about his doing X, then 'stepping back' can have no psychological content implied by it. For suppose A arrives inductively at the conclusion that he will do X, just as a prediction, and then says 'I see: I shall be

doing *X* at such-and-such a time; please take that into account in your arrangements'. By Hampshire's argument he *must* be here announcing a decision; and if 'stepping back' is used in the way just described, it will logically follow, or be entailed, that he must have 'stepped back' from the prediction. But evidently it does *not* logically follow from these premisses that he had any psychological experience of deciding, or reviewing the situation and choosing to acquiesce in it, at all. On this interpretation, 'stepping back' will be unavoidable, but also, quite probably, unnoticeable.

Alternatively, 'stepping back' may be taken to have some psychological content, of reviewing the situation in which the prediction has been given, deciding whether to acquiesce, etc. In this case stepping back will not be unnoticeable, but will also not be unavoidable; since (by the previous argument) it may, as a matter of psychological fact, just not happen. But perhaps Hampshire could say that there is still a sense in which 'stepping back' is unavoidable, even if the notion is given psychological content. For there does seem to be a sense in which it is unavoidable, once it has *occurred* to a man that there is something here to be decided, and that if he goes along with the prediction, he will indeed be acquiescing in it.

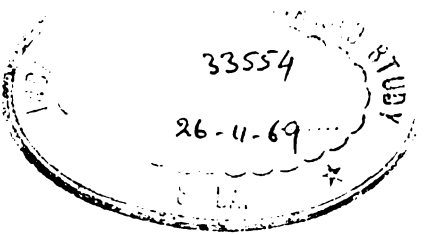
Here, however, Miss Murdoch urges a second difficulty: that, on these premisses, there is now a matter of moral choice. For granted (what is perhaps Hampshire's own way of putting it) that an agent always *can* step back in the more substantial, psychological, sense, and decide whether to acquiesce

or not, might he not have a general policy against doing so, part, perhaps, of some world outlook? If Hampshire insists on the unavoidability of actual stepping back, it looks as though he is really making a moral point: it is not so much stepping back that we cannot avoid, as the responsibility which follows from our being able to step back. To this, Hampshire's reply seems to be that, in a sense, he agrees; but that the moral objection to refusing to step back in the way Miss Murdoch describes, is just the moral objection to any sort of evasion (p. 88), or, one might say, self-deceit. The sort of policy Miss Murdoch refers to does indeed, if Hampshire's premisses are granted, just sound like a sustained attempt to convince oneself that certain things are unalterable necessities of one's circumstances when they are not.

Miss Murdoch has a more general criticism on these lines, that Hampshire's views depend on an evaluatively motivated picture of the mind, which sharply distinguishes between 'reason' and 'will'. To this, again, Hampshire is prepared to agree, in the sense at least that he does distinguish decision and prediction, practical and theoretical knowledge, and that he does not regard the classification of the powers of the mind on which this depends as final, any more than any other (see his closing remarks in the discussion). However, I think this concession applies most directly to the parts of his views which are the centre of the discussion. Taken more generally, Miss Murdoch's point needs some qualification. Where Hampshire is concerned with the degrees of freedom, and the increase of freedom, he stresses the

connection of this with knowledge, self-awareness, and the ability to think of alternate courses of action; and this emphasis precisely brings to the fore the rôle of reason and thought in free action, as opposed to the tradition which has stressed only the rôle of the will. In Hampshire's theory, as in fact, there is more to freedom than the freedom of the will.

THE END



PRINTED IN GREAT BRITAIN BY
LOWE AND BRYDONE (PRINTERS) LIMITED, LONDON